UNDP LAC WORKING PAPER SERIES

BACKGROUND PAPER OF THE REGIONAL HUMAN DEVELOPMENT REPORT 2021

24

UNDP LAC
WORKING PAPER SERIES

TIME FOR A NEW COURSE: AN ESSAY ON SOCIAL PROTECTION AND GROWTH IN LATIN AMERICA

JUNE 2021

Santiago Levy
Brookings Institution and UNDP
Guillermo Cruces
CEDLAS Universidad Nacional de La Plata

TIME FOR A NEW COURSE: AN ESSAY ON SOCIAL PROTECTION AND GROWTH IN LATIN AMERICA[1]

Santiago Levy[2]
Guillermo Cruces[3]

Abstract

We develop a framework to analyze the effects of social protection policies on two dimensions: their efficacy to protect households against risks, reduce poverty and mitigate inequality; and their impact on productivity and long-term growth. Our framework pays close attention to the legal and institutional context in which these policies are deployed, and to their interaction with other policies that also impact the social and economic outcomes that we focus on, particularly those related to taxation and the performance of markets. We argue that social protection policies provide incomplete and erratic protection against risks, do not redistribute enough and hurt productivity—in sum, that they are inconducive to socially inclusive growth. We propose a new social protection architecture based on the principle of universality with respect to the relevant population and discuss its implications for health; retirement pensions; protections against disability, death and the loss of employment; poverty; inequality; and for firm and worker behavior among various margins that impact productivity. We discuss the tax implications of this architecture. We argue that, jointly considered, our proposals offer a route to fiscally sustainable and socially inclusive growth.

JEL Code: D63, H55, I13, I14, I32, J24, O4
Keywords: Social insurance, poverty, inequality, productivity, universality, socially inclusive growth

Background Paper for the UNDP LAC Regional Human Development Report 2021. The views and recommendations presented in this paper are from the authors and do not necessarily represent the official position of UNDP.

[1] We express our gratitude to Luis Felipe López-Calva and Marcela Meléndez, the Bureau's director and chief economist, respectively, for inviting us to work in the Regional Human Development Report, and for their very useful suggestions and comments. The document draws evidence from nine country studies prepared for the same report: Argentina (Alzua and Pacheco), Brazil (Firpo and Portella), Chile (Morales and Olate), Colombia (Alvarado, Meléndez and Pantoja), the Dominican Republic (Barinas and Ñopo), Ecuador (Ñopo and Peña), Honduras (Ham and Membreño), Mexico (Correa, García de Alba, Martínez, Martínez, Mandoza and Sulmont) and Peru (Ñopo). We want to express our gratitude to the authors of the country studies for very useful interactions and discussions during the preparation of this document. In parallel, Alejandro Bonvecchi, Augusto de la Torre, Leonardo Gasparini, Alain Ize and Michael Walton also provided very useful comments and suggestions, and we also express our gratitude to them.
[2] Senior Fellow and Consultant. Brookings Institution and UNDP. United States. santiagolevy4@gmail.com
[3] Researcher and Deputy Director of the Center for Distributive, Labor and Social Studies (CEDLAS) Facultad de Ciencias Económicas at Universidad Nacional de La Plata (UNLP) and CONICET. Argentina. gcruces@cedlas.org

Recommended citation: Levy, S. and Cruces, G. (2021). Time for a New Course: An Essay on Social Protection and Growth in Latin America. UNDP LAC Working Paper No. 24. Background Paper for the UNDP LAC 2021 Regional Human Development Report.

Contents

1. INTRODUCTION 5

2. CONCEPTUAL FRAMEWORK 10
2.1 Laws and Institutions 10
2.1.1. The realm of entrepreneur-worker relations 11
2.1.2. The realm of taxes and transfers 15
2.1.3. The realm of market conditions 16
2.1.4. Institutions and social and economic outcomes 17

2.2 Social Protection Policies 18
2.2.1. Identifying social protection policies 18
2.2.2. Formality and informality: workers 22
2.2.3. Formality and informality: firms 23
2.2.4. Formality and informality: legality and productivity 23
2.2.5. Labor market outcomes and social protection 25
2.2.6. Firms and social protection 25

2.3 Social Protection Outcomes and Productivity Outcomes: A Unified Process 27
2.3.1. Firms and aggregate productivity 27
2.3.2. The composition of employment 28
2.3.3. The size distribution of firms 29
2.3.4. The productivity of firms 30
2.3.5. Simultaneous determination of social protection and productivity outcomes 32

3. SHORTCOMINGS OF SOCIAL PROTECTION 34
3.1 The Formal-Informal Composition of the Economy 37
3.1.1. Formal-informal transitions 39
3.1.2. Formal-informal transitions by income level 39
3.1.3. Overlap between the formal and informal earnings distributions 39

3.2 Social Protection Outcomes: Insurance 40
3.2.1. Limited coverage 40
3.2.2. Erratic coverage 41

3.3 Social Protection Outcomes: Poverty 43
3.3.1. Income transfers 45
3.3.2. Income transfers versus insurance 46
3.3.3. Interphase between insurance, minimum wages and poverty programs 47
3.3.4. Permanent versus transitory transfers 49

3.4 Social Protection Outcomes: Inequality 50
3.4.1. Contributory social insurance and inequality: incidence of contributions 51
3.4.2. Contributory social insurance and inequality: cross-subsidies 54
3.4.3. Non-contributory social insurance and inequality 58
3.4.4. Summing up on social insurance 59
3.4.5. Other taxes and transfers and inequality 60
3.4.6. Minimum wages and inequality: income floors for workers 62
3.4.7. Minimum wages and inequality: thresholds 65

3.5 Social Protection Policies and Productivity 67
3.5.1. The tax on formality and the subsidy to informality 67
3.5.2. Special regimes for the self-employed and micro and small firms 71
3.5.3. The size of firms 74
3.5.4. Digression: imperfect contract enforcement and firm size 76

3.5.5. Interactions between size and formal-informal composition 77
3.5.6. Social protection policies, resource allocation and productivity 79
3.5.7. Formalization programs and productivity 80
3.5.8. The misuse of human capital 81

3.6 Trade-offs Between Social Protection and Productivity Objectives 82
3.6.1. Coverage of insurance 82
3.6.2. Poverty 83
3.6.3. Inequality 84

3.7 Social Protection and the Strength of the Rule of Law 86

3.8 Social Protection and Growth 90
3.8.1. Additional factors affecting growth 90
3.8.2. The region's growth strategy 91
3.8.3. Interphase between the region's growth and social protection strategies 96
3.8.4. Growth versus socially inclusive growth 99

4. HOW SHOULD SOCIAL PROTECTION BE STRUCTURED? 102
4.1 Unbundling Risks and Insurance 103

4.2 Illness 104

4.3 Old-Age Poverty and Longevity 106
4.3.1. Old-age poverty and universal non-contributory pensions 106
4.3.2. Consumption smoothing and universal contributory pensions 107
4.3.3. Considering jointly non-contributory and contributory pensions 111

4.4 Death and Disability 113

4.5 Accidents at Work 113

4.6 Separation from Firms 114
4.6.1. Cause free termination payments 115
4.6.2. Enhanced unemployment insurance and dismissal compensations 116

4.7 Universal Social Insurance 118

4.8 Universal Social Insurance and Productivity 120

4.9 Complementary Policies Against Poverty 122
4.9.1. Equal social insurance for poor and non-poor households 122
4.9.2. Improving labor markets for poor workers 124
4.9.3. Complementary income transfers for the poor 124

4.10 Universal Basic Income 126
4.10.1. UBI as permanent insurance 126
4.10.2. UBI as contingent insurance 127
4.10.3. UBI versus targeted programs to transfer income to the poor 128

4.11 Taxation and Social Protection 130
4.11.1. Identifying the fiscal effort 131
4.11.2. Identifying sources of revenue 133
4.11.3. Taxation, informality and productivity 137
4.11.4. Further observations on taxation and social protection 139

4.12 Income Inequality 140
4.12.1. Measuring the impact of tax and social protection policies on inequality 140
4.12.2. Minimum wages and income inequality 143

4.13 Universal Social Protection and the Strength of the Rule of Law 145
4.13.1. Better laws 145
4.13.2. Better enforcement of the laws 146
4.13.3. The end of informality? 146

4.14 Towards More Encompassing Social Protection 147

5. UNIVERSAL SOCIAL PROTECTION IS POSSIBLE IN LATIN AMERICA 149
5.1 Looking Back 149

5.2 Looking Forward 151

5.3 A Complicated but Not Impossible Political Economy 152
5.3.1. Taxation 152
5.3.2. Special interest groups 154

5.4 The Centrality of Ideas 156

REFERENCES 159

Acronyms

CCT = conditional cash transfers
CSI = contributory social insurance
EUI = enhanced unemployment insurance
NCSI = non-contributory social insurance
OECD = Organization for Economic Co-operation and Development
PAYG = pay-as-you-go
UBI = universal basic income
UHI = universal health insurance
UI = unemployment insurance
USI = universal social insurance

Symbols

$L(L_{CSI}, L_{JS}, L_{MW}, L_{NCSI}, L_{ENF}) = L(.)$ = realm of entrepreneur-worker relations
L_{CSI} = regulations on contributory social insurance
L_{JS} = regulations on job stability
L_{MW} = regulations on minimum wages
L_{NCSI} = regulations on non-contributory social insurance
L_{ENF} = institutions enforcing ($L_{CSI} + L_{JS} + L_{MW}$)

$T(T_{PIT}, T_{CIT}, T_{VAT}, T_{OTH}, T_{TRN}, T_{ENF}) = T(.)$ = realm of taxes and transfers
T_{PIT} = personal income taxes
T_{CIT} = corporate income taxes
T_{VAT} = consumption taxes
T_{OTH} = other taxes (property, environment)
T_{TRN} = transfers to households
T_{ENF} = institutions enforcing taxes

$M(M_{FTI}, M_{COM}, M_{PDP}, M_{ENF}) = M(.)$ = realm of market conditions
M_{FTI} = policies on foreign trade and investment
M_{COM} = policies on domestic competition
M_{PDP} = policies on productive development
M_{ENF} = institutions enforcing credit and commercial contracts

$\mathbf{E}\{ L(.), T(.), M(.) \}$ = the environment

1. Introduction

After the lost decade of the 1980s, under the broad rubric of social policy, Latin America increased social spending to protect workers against risks, alleviate poverty and mitigate inequalities. In parallel to long-standing programs, new ones were created to expand the coverage of pensions, health, child allowances and childcare services and other benefits; to redistribute in favor of low-income households through direct transfers, minimum wages, subsidies for basic consumption goods and energy; and to help micro and small firms through favorable tax treatments and subsidized credits. In parallel, large resources were devoted to investments in human capital, from programs promoting early child development, to programs to expand the coverage and improve the quality of education, to training programs to increase the skills and abilities of individuals participating in the labor force.

But three decades later, despite these valuable efforts, inequality is still high; and while poverty has fallen, in most countries it is above what would be expected given their income levels. Moreover, the majority of workers in the region, aside from receiving incomplete and erratic protection against risks, lack jobs where they can develop their potential and increase their incomes overtime; indeed, despite advances in schooling, too many are still self-employed in precarious jobs, or employed by fragile and small low-productivity firms. Although social indicators have improved, in most Latin American societies precariousness and exclusion are still the norm rather than the exception.

Why have results disappointed? Is it because social spending is insufficient or because the social policies that are promoted are flawed? Or are deficiencies in other domains of public policy impeding the region from fully reaping the benefits from its social investments?

Social policy encompasses many initiatives. In this essay we divide them into those focused on protecting households against risks, combating poverty and mitigating income inequalities; and those aimed at augmenting individuals' schooling, skills and abilities. Henceforth, we denote the first as social protection policies and the second as policies to invest in human capital. Although both are equally important, here we only focus on the first.

We have witnessed substantive advances in our understanding of social protection in Latin America. Better and more abundant data combined with improved statistical techniques have resulted in the accumulation of a large and continuously growing stock of evidence about the impact of individual policies. This valuable knowledge has helped policymakers adjust existing policies and improve the design of new ones. Despite these advances, it is our belief that an overall view of social protection policies and programs has been lacking. Our understanding of how individual programs function on their own is not matched by our understanding of how they interact with each other. Nor is it matched by an understanding of how these programs—considered *in toto*—interact with other public policies implemented in the region to stimulate growth, jointly resulting in a shared and sustainable prosperity.

At the end of the day, however, social protection outcomes do not depend only on individual social protection policies or programs—they depend on the joint effects of all of them. Clearly identifying the objectives and target populations of each program is not enough, nor is the fact that some programs have positive evaluations. Policies and programs need a comprehensive logic to ensure that they work harmoniously together and complement each other. Social protection should not be an agglomeration of many programs, the more the merrier. It should be a coherent set of policies and programs that are consistent with each other, and that jointly contribute to improved social outcomes.

At the same time, consideration of the contexts in which these policies are deployed is also essential to understand why they work or fail. Clearly, it makes a large difference whether they are implemented in a context where growth is vigorous or anemic. Growth depends on a multitude of factors, of course; but all in all, and unlike policies to invest in human capital, we consider that the impact of social protection policies on growth has not received the recognition that it deserves. This is particularly important in Latin America, the slowest-growing region in the world over the last two decades.

This is a serious omission because social protection policies have a direct bearing on key determinants of productivity and long-term growth: the division of the workforce between those that manage firms, those that work for firms and those that work for themselves; the size of firms and the technologies they adopt; the legal or illegal nature of contracts between firms and workers and the short- or long-run nature of their associations; workers' incentives to invest in schooling and opportunities to acquire skills and learn on the job; firms' incentives to invest in worker training; and the matches between firms of various sizes and productivities with workers of different skills.

In parallel, it is also clear that the strength of the rule of law—understood as the degree of compliance with a country's laws and regulations given citizens' predispositions to observe them, and the effectiveness of institutions enforcing them—matters for social protection. Under strong rule of law, the obligations to contribute to social insurance are enforced and minimum wages are observed, to cite just two examples. Clearly, when the rule of law is weak, social protection outcomes are hurt.

Many factors determine the strength of the rule of law. But, as with growth, we consider that there is insufficient recognition that social protection laws and policies are one of them. When these laws are well designed and function correctly, they strengthen the bond between individuals and the state, increase social capital and contribute to a general attitude that favors respect and compliance with all laws. But when this is not so, as often happens in Latin America, they can undermine the rule of law. In this region, social protection laws at times are expressions of political *desiderata*, rather than guides for behavior. At other times, their complexity and ambiguity make them difficult to comply with and enforce, inducing their replacement by tacitly accepted social norms, thereby weakening the rule of law.

Prosperous and equitable societies cannot be constructed in contexts of slow growth or stagnation and uncertain legality. The efforts made through social protection policies and those made through other public policies need to complement one another, and this complementarity needs to consider the two-way interactions between them.

These perspectives motivate this essay. We are interested in how different social protection policies, particularly protections against risks, inequality and poverty, interact with each other and impact social outcomes; and we are interested in how these policies also impact productivity and long-term growth. The essay is an attempt to look at the forest, rather than at individual trees. Of course, a detailed understanding of individual trees is very valuable; and we interpret much of the research on social protection in the region as doing just that. But a view of how one tree interacts with other trees, and how the forest in turn impacts the environment, is essential, as is a view of how the environment also impacts the performance of the forest.

A holistic analysis of social protection sheds light on the questions posed above. The region's underwhelming social protection outcomes may partly result from insufficient resources but may also partly result from the fact that the policies and programs deployed—jointly considered—may themselves be obstructing better outcomes.

However, a holistic analysis is not without problems. First and foremost, as the number of policies considered increases, the interactions between them grow exponentially. In this context, careful identification of causality is not possible. As a substitute, we offer a hopefully clear conceptual framework complemented with some of the available empirical evidence. On the other hand, because we discuss many countries at the same time, we refer to the main features of social protection in the region, but do not provide a detailed analysis of any single country. There is a trade-off between scope and depth, and in our choice for scope we have inevitably sacrificed some depth.

With those caveats in mind, we make the case that, independently of the need for more social spending, shortcomings in the design of social protection policies go a long way to explaining disappointing social protection outcomes. We also make the case that these shortcomings have partly offset efforts made through other public policies to increase productivity and strengthen the rule of law. More precisely, we argue that social protection policies fail in three distinct but related dimensions. They:

- do not redistribute enough nor protect workers effectively against risks;
- depress productivity and hurt long-term growth; and,
- at times, undermine the rule of law.

On this basis, we argue that a change in course is necessary. Social protection outcomes will not improve only by increasing spending, either by expanding existing programs or adding new ones. There are serious flaws in the region's social protection architecture that need attention.

Before Covid-19 hit Latin America, changing social protection was extremely important, but after the crisis associated with the pandemic, the task has become urgent for three reasons. First, a vigorous increase in social spending is indispensable to reverse the increases in poverty and inequality caused by the pandemic. Second, deteriorated fiscal conditions across the region require more taxes. Third, countries need to strengthen their efforts to overcome the growth-depressing effects of the pandemic.

A central message of this essay is that, as things stand today, increasing taxes to finance higher social spending will not deliver more social inclusion, and will likely offset other efforts made to resume growth. Those concerned with strengthening fiscal sustainability and accelerating growth need to pay careful attention to the structure of social protection if they want to reach their aims; and those concerned with poverty, inequality and social welfare need to pay careful attention to the impact of social protection on growth and the fiscal stance, if they want to achieve theirs. Unless both groups do so, it is difficult to imagine that the region can recover from the damage caused by Covid-19 and embark on socially inclusive growth.

The essay has four parts. Chapter two introduces key concepts and provides a framework to place social protection policies within the broader set of public policies that determine social and economic outcomes. This framework serves to make two points that permeate the essay. First, social protection outcomes depend on all policies, not only social protection ones; second, social protection policies influence other outcomes, not only social ones. Policies in the domain of taxation and market regulation (critical determinants of resource allocation and productivity) impact social protection; and social protection policies (critical determinants of social welfare) impact resource allocation and productivity.

Chapter three is a retrospective analysis of social protection in Latin America, with emphasis on its main shortcomings. Importantly, we do not provide an assessment of all social protection policies. There are individual successes that we do not dwell on, focusing instead on those aspects that we consider to be more problematic. We discuss the impact of these policies on insurance, poverty and income inequality. We also consider their impact on resource allocation and productivity. (As an aside, we hope the analysis serves as a bridge between the literature on social protection and the literature on long-term growth and motivates research on their interconnections.)

Our overall conclusion is that the current social protection architecture—its laws, institutions and programs—is an impediment to socially inclusive growth in the region. More resources are certainly needed to expand insurance coverage, fight poverty and reduce inequalities. But this is not nearly enough. If we want social protection be more effective and inclusive, fiscally sustainable and growth-enhancing, fundamental laws and institutions need to change.

Chapter four is normative, presenting a discussion of how social protection should be structured in Latin America. The central guiding principle is universality with respect to the relevant population; a principle that has large implications for social inclusion, social entitlements and the equality of individuals *vis-à-vis* the law. We argue that this principle contributes to better social and economic outcomes, and we make proposals to implement it in health, various types of pensions, protections against the loss of employment and policies to combat poverty.

Many suggestions are constantly put forth to improve social protection in the region: expand transfers for the poor, raise minimum wages, introduce a universal basic income, reform this or that health or pension program, create unemployment insurance and so on. We consider that the principle of universality with respect to the relevant population provides a useful guide as to which policies need to be pursued and how they should be financed. Of course, a principle is not a rigid blueprint. Many variants and combinations of policies are consistent with this principle, whose pros and cons depend on the characteristics of individual countries, including political preferences for taxation and social solidarity. Nevertheless, if social protection is going to be more effective than in the past, it needs to be more than a collection of programs; it needs guiding principles, and its various pieces need to form a coherent whole.

Restructuring social protection is impossible without more taxes. In this vein, chapter four presents rough estimates of the fiscal costs of our proposals and equally rough identifications of sources of revenues to make them fiscally sustainable. But that is not enough. Taxes also impact productivity. Thus, it is necessary to ensure that the incentives associated with social protection and those associated with taxation jointly point in the direction of higher productivity. Of course, taxes also impact the distribution of income, making it necessary to ensure that, together with social protection programs, they have the desired effect on inequality. In this context, the chapter makes the key point that not just any combination of taxes and contributions will do, even if it is sufficient to finance social protection programs. The region needs to stop taxing the high-productivity sectors of the economy and subsidizing the low-productivity ones as a vehicle to reduce income inequalities; rather, it needs to directly tax high-income individuals and subsidize low-income ones without distorting occupational choices and firm behavior.

Chapter five concludes with some brief remarks on the interplay of interest and ideas in social protection. We recognize the complex political economy of reforming social protection and make some general observations, but do not provide a careful assessment. That said, the chapter argues that restructuring social protection is impossible without a willingness

to reassess long-standing views and ideas, adapting them to the realities of Latin America. The region's social protection efforts have been under-rewarded partly because some ideas continue to dominate the debate and frame policies despite the fact that they have not worked for many decades. New views and ideas are needed. Of course, this is not to say that the views expressed here are the right ones; some may not be, and all can surely be improved. Nor is it to say that interests are irrelevant. But it is to say, remembering Keynes, that policymakers may be more imprisoned by ideas than by interests, and that in the end it is ideas that matter most.

Change in social protection can occur in many forms, depending on countries' political contexts. But change is essential for the region to escape from its present predicament. Hopefully this essay contributes to this change.

2. Conceptual Framework

2.1. Laws and Institutions[1]

Social and economic outcomes in every country are shaped by its laws and institutions; these set the rules of the game that individuals face when deciding to start a business, produce, invest, take risks (innovate, lend, borrow) or to work for themselves or for somebody else. In this essay we label all these laws, institutions and associated policies as "the environment", represented by $\mathbf{E}$. The environment determines the efficiency with which economic activity is undertaken, but also the distribution of its fruits among the individuals participating in it, and the degree to which they are sheltered from various contingencies. The environment shapes the social fabric and the degree of solidarity between members of society.

Although all encompassing, it is useful to think of $\mathbf{E}$ as made up of three broad areas: the institutions, laws and policies that govern relations between firms and workers; the ones that determine taxation and transfers; and those that regulate the functioning of markets. In this essay we pay careful attention to the first because, for reasons that will soon be apparent, they are central to social protection. We also pay some attention to the second, although we do not discuss all aspects of taxation. Institutions associated with the functioning of markets are also important for social protection, although we only address them when they are central to a particular issue (such as their impact on firm size or the relationship between social protection and growth).

Despite similarities, Latin America is a heterogeneous region, and laws and institutions vary across countries. Many are specific to individual countries, and it is impossible to discuss each. To analyze social protection throughout the region, the discussion must inevitably be somewhat abstract. As a result, in this essay we refer to "a country" which, *mutatis mutandis*, may be Argentina, Brazil, Ecuador, Peru or any other. Here we focus on laws and policies that are relevant to all.

[1] This section draws heavily from Levy (2018).

2.1.1. The realm of entrepreneur-worker relations

The institutions and laws governing the modalities under which entrepreneurs and workers can associate to form firms and establishing their respective rights and obligations, as well as the institutions in charge of enforcing these rights and obligations, are all gathered in the "realm of entrepreneur-worker relations." We represent it by L(..., L_i, ...), where L_i in turn refers to the laws, policies and programs over a particular dimension of these relations. Although these dimensions are numerous, we list five that we consider to be the most relevant in Latin America:

- L_{CSI}: laws determining the universe of firms and workers mandated to participate in contributory social insurance programs, where taxes proportional to workers' earnings are earmarked to finance benefits like health, disability, life and unemployment insurance, and retirement pensions;
- L_{JS}: laws on job stability regulating the conditions under which firms can dismiss workers and the compensations that must be paid in each case (separate from unemployment insurance considered in L_{CSI});
- L_{MW}: laws on minimum wages;
- L_{NCSI}: programs providing health services and retirement pensions, among others, for workers excluded from L_{CSI}. These programs, financed from general revenues, are labelled non-contributory social insurance; and,
- L_{ENF}: labor tribunals adjudicating disputes on L_{JS} and institutions collecting contributions for L_{CSI} and enforcing minimum wage laws L_{MW}.

In sum, L(.) = (L_{CSI}, L_{JS}, ..., L_{ENF}) is an abstract representation of the immensely complex set of laws, institutions and practices that regulate how firms and workers in Latin America interact, including the social benefits that workers derive from those interactions. This representation is incomplete as it excludes dimensions like regulations on child labor or unions. But the ones considered here are sufficient for our purposes, as they capture core policies to insure workers against the risks of illness, disability, longevity, death and loss of employment.

Each dimension of L(.) encompasses many institutions, laws and policies. Take, for instance, L_{CSI}. In some countries, all contributory social insurance programs are administered by specialized social security institutes. In others, however, they are managed by different agencies, for example one for health and another one for pensions. In some cases, these institutes are only charged with financing, but in others they also provide services; in other cases, services are provided by subnational governments or subcontracted to private providers. Practices vary, but it is not necessary at this point to provide details; suffice it to say that L_{CSI} is a symbolic representation of the laws and institutions associated with contributory social insurance.

Similarly, L_{ENF} also encompasses many institutions, in this case across different branches of government. Adjudication of labor disputes is generally carried out by tribunals in the judicial branch, while institutions in charge of labor inspections are usually in labor ministries of the executive branch. In turn, enforcement of contributions may be carried out by social security institutes themselves, or by specialized agencies of the ministries of finance charged with tax collections. Practices vary and at this point all that is necessary is to visualize L_{ENF} as representing the efficacy with which the laws associated with L_{CSI}, L_{JS} and L_{MW} are enforced.

Critically, some dimensions of L(.) are interdependent. In all countries there is a legal connection between L_{CSI}, L_{JS} and L_{MW}. In discussions on social protection, L_{CSI} is considered as

a separate policy from L_{JS} and L_{MW}, the former under the label of contributory social insurance and the latter under the label of labor regulations. But this is a mistake. Regardless of labels, these three dimensions are legally joined: workers in firms covered by L_{CSI} are always covered by L_{JS} and L_{MW}, and firms are required to comply with all three at the same time.

Indeed, ($L_{CSI} + L_{JS} + L_{MW}$) needs to be conceptualized as a single regime—the central mechanism by which Latin American societies aspire to insure their workers against key risks, provide them with a minimum income and, through provisions on job stability, ensure that they are permanently protected.

In this essay we will sometimes refer to these three dimensions as the "($L_{CSI} + L_{JS} + L_{MW}$) regime" or, alternatively, as "($L_{CSI} + L_{JS} + L_{MW}$) laws". It is difficult to exaggerate the importance of this regime. Contributory social insurance, minimum wages and labor regulations *in toto* are the result of an effort that began in Latin America over 75 years ago and that, to this day, is expected to be the centerpiece of its social protection. This regime has affected views and ideas on social protection in more ways than is often recognized and despite some changes since its inception, its basic structure still stands.

But this regime does not protect all workers. Non-remunerated workers are excluded in all countries, and rural or domestic workers are excluded in some. In others, only workers in a "relation of dependency and subordination" with a firm are included, a relation that occurs when workers are paid a periodic monetary remuneration labelled wage or salary in exchange for performing the tasks set by firms. Workers in this relation are at times called dependent and at times salaried, and their rights and obligations with respect to social insurance and labor regulations are very different from the rest, usually called non-dependent or non-salaried workers.

Non-dependent workers participate in the labor market in multiple forms. Very importantly, some can be in firms. For instance, they may form enterprises like cooperatives or joint ventures and share the surplus produced according to cultural norms, rather than through periodic monetary remunerations like wages or salaries. This is particularly relevant when members of an enterprise are related to one another and constitute a family firm. In these cases, there is no relation of dependency and subordination and no salaries or wages, and workers are excluded from the ($L_{CSI} + L_{JS} + L_{MW}$) regime. Put differently, in Latin America not all workers in firms are dependent workers.

Further, when firms are very small, with two or three workers, it is not easy to establish the existence of a relation of dependency and subordination since a sharp line cannot be drawn between situations where workers join efforts in a cooperative enterprise without a boss giving orders, and situations where *de facto* there is a manager-entrepreneur, with the rest of the workers subordinated to them, even if there is no explicit recognition that there is a boss and no written contract specifying who does what when. Because the interpretation of the relation of dependency and subordination can be ambiguous, some firms live in the borderline of having to comply with the ($L_{CSI} + L_{JS} + L_{MW}$) regime, and many do not, leaving their workers without its protections. Enforcement through the institutions in L_{ENF} is ineffective because it is not clear that ($L_{CSI} + L_{JS} + L_{MW}$) laws apply.

In some countries in the region, self-employed workers—who by definition are not in a relation of dependency and subordination with a firm—are excluded from the ($L_{CSI} + L_{JS} + L_{MW}$)

regime. In others, only those with incomes above a certain threshold are included (although in their case regulations on job stability and minimum wages do not apply).[2]

Self-employment is very important in Latin America, and it is indispensable to probe a bit more into its nature. A self-employed worker is usually thought of as a single individual working completely on their own: a gardener, plumber or electrician providing services from home to home; a person shining shoes on the streets or cleaning windshields while cars are stopped at a red light; a peasant tilling their own plot of land; or a lawyer or a doctor working from home without any assistance.

But self-employment can be something else. In many countries, economic activity takes place in enterprises that are not incorporated as a legal entity. In these cases, the assets of the enterprise are undistinguishable from the assets of the owner. These enterprises sometimes carry out their activities in the streets or inside the household, and produce goods (food, clothing, furniture) or services (transportation, cleaning) with the participation of various individuals. Their size, as measured by the number of workers, is usually very small, but fluid. If demand is low, perhaps only two individuals are involved, but this can rapidly increase to three or four if demand is high. A worker-owner may be selling to final consumers, thus looking like a self-employed individual in a narrow sense, but behind them at home, or next to them in the street stand, other workers help him.

Self-employed workers are at times referred to as independent workers, a terminology that reflects the fact that they may manage a very small enterprises, not that they necessarily work completely on their own. But regardless of how they are labelled, the key point is that oftentimes they are not working alone. Other workers are involved. *De facto*, a firm is there. But because the relation of dependency and subordination is blurred in this firm, it may exist outside the domain of the (L_{CSI} + L_{JS} + L_{MW}) regime.

The message we want to convey here is that the border between a truly self-employed worker and a self-employed worker-cum-very small firm is fuzzy. The firms of the self-employed may not be registered as such with the relevant authorities, but that should not detract from the critical fact that, from an economic point of view, they are firms: units of production where two or more individuals combine their efforts with capital and intermediate inputs to produce output for sale in the market. These firms, usually labelled micro firms, are critical in Latin America for two reasons. First, all the workers involved, including the owner-manager, may be excluded from the (L_{CSI} + L_{JS} + L_{MW}) regime. Second, together with the truly self-employed, they account for around 50 percent of the labor force, sometimes more.

There is another category of workers excluded from the (L_{CSI} + L_{JS} + L_{MW}) regime that is important in the region: those without monetary remuneration. In Ecuador they account for 13 percent of the labor force, in Honduras, 12 percent and in Peru, 10 percent. Even in Mexico, a country with a higher per capita income, 4.3 percent of workers fall in this category. Clearly, if

[2] The Dominican Republic, Ecuador, Mexico and Peru, among others, exclude all self-employed workers. Chile requires them to contribute if their earnings exceed four minimum wages. Colombia does too but sets the threshold at one minimum wage (although the minimum wage in Colombia is set at the 50th percentile of the wage distribution versus the 16th percentile in Chile). Honduras also requires them to contribute, but workers are exempt if they claim that they cannot comply. Argentina and Brazil, on the other hand, require that all self-employed workers contribute. All references to countries' laws, data, tables and graphs not expressly referenced in the document come from the nine country studies that form part of the UNDP Report.

"without monetary remuneration" were interpreted literally, these workers would not survive. Rather, they are employed in family firms, usually relatives of the firm owner, paid in-kind, living and sharing in the household running the firm.

Excluding workers from (L_{CSI} + L_{JS} + L_{MW}) laws has a very important implication for social protection in Latin America: even if these laws were fully observed, their coverage would be far from universal. However, they are not always observed. Because the L_{ENF} institutions charged with enforcing them work imperfectly, many firms evade these laws, particularly small ones. Evasion of these laws further enlarges the set of workers excluded from the coverage of the (L_{CSI} + L_{JS} + L_{MW}) regime.

In sum, explicit exemptions, ambiguous obligations and outright non-compliance jointly result in a light- to dark-grey zone between legal exclusion and illegal evasion of (L_{CSI} + L_{JS} + L_{MW}) laws. This grey zone matters greatly for social protection because it limits workers' rights to social insurance. Some workers are entitled to the protections derived from these laws, some are not and some others, well, maybe.

This is the critical flaw of the (L_{CSI} + L_{JS} + L_{MW}) regime—one that, as we discuss in this essay, has plagued social protection in the region for a very long time.

Incomplete coverage of (L_{CSI} + L_{JS} + L_{MW}) laws explains the existence of L_{NCSI}. An observer from a European country looking at Latin America would be very familiar with these laws, in large measure because Latin America imported them from Europe. That observer would also not be surprised by the fact that these laws are imperfectly enforced, because they are nowhere perfectly enforced (although they would be surprised at the extent of evasion). But the existence of a parallel system of social insurance funded from general tax revenues for workers who, for whatever reason, are excluded from L_{CSI} would in all likelihood call their attention, all the more so if, as we demonstrate below, their earnings are similar and, in some cases, higher than the earning of those covered by the (L_{CSI} + L_{JS} + L_{MW}) regime.

Thus, despite some apparent similarities, the "realm of entrepreneur-worker relations" in Latin America is different from that of most OECD countries. In addition to exclusions, ambiguities and weaker enforcement of (L_{CSI} + L_{JS} + L_{MW}) laws, and the presence of L_{NCSI} programs, other important differences are:

- programs in L_{CSI} are bundled together and firms and workers are obligated to contribute to all at the same time. These programs provide insurance against risks for those who contribute, but also often include items unrelated to risks like family allowances (Argentina), housing (Mexico) or childcare (Dominican Republic and Mexico), or benefits for the whole population, like education (Brazil), early child development (Colombia) and training (Peru);
- with a few exceptions like Argentina and Chile, the absence of unemployment insurance from L_{CSI};
- low workers' valuation of L_{CSI} programs because of the fact that they are bundled together, some do not benefit them, the quality of services like health or childcare is unsatisfactory, and deficiencies in the rules to access pensions, among many factors elaborated on below; and,

- low workers' valuation of L_{JS} regulations because transactions and legal costs are high relative to the expected benefits given that labor tribunals are sometimes corrupt or take a long time to adjudicate disputes.

2.1.2. The realm of taxes and transfers

The second area that makes up $\mathbf{E}$ is labelled the "realm of taxes and transfers," represented by $T(T_{PIT}, T_{CIT}, ..., T_{ENF})$. This is also a very broad area, but for our purposes we highlight six dimensions:

- T_{PIT}: laws on personal income taxation, determining the universe of individuals required to comply and the progressivity of the tax schedule;
- T_{CIT}: laws on corporate income taxation including, very importantly for the region, special regimes for micro or small firms;
- T_{VAT}: laws taxing consumption, under the form of a VAT or sales or excise taxes;
- T_{OTH}: laws taxing real estate, environmental externalities, financial transactions, energy (which may be subsidized) or other goods, services or activities;
- T_{TRN}: transfers to households not contemplated in L_{CSI} or L_{NCSI}, which may be in-kind (like childcare or health services) or in cash, and may be universal or targeted to a subset of households through some form of means testing; and,
- T_{ENF}: institutions in charge of enforcing taxes and delivering transfers.

As with L(.), each dimension of T(.) may be associated with various laws and institutions. Take, for instance, T_{OTH}. Real estate taxes are usually set and collected by subnational governments, while taxes on energy are managed by national ones. But these specificities are not relevant for our purposes; it is sufficient to know that T_{OTH} captures all other taxes except for those on income or consumption. Similarly, T_{TRN} involves many institutions, depending on the nature of the transfer; but again, this is not relevant at this point. For our purposes, what matters is that households receive transfers from the government (whose rules we explore below).

In principle, contributions for L_{CSI} on one hand, and taxes in T_{PIT} and T_{CIT} on the other, are distinct. The former generates revenues specifically earmarked for benefits to the workers paying them; the latter generate revenues that can be freely disposed by the government. But in Latin America, they sometimes overlap. As already noted, in some countries L_{CSI} contributions are also used to finance benefits for the whole population. They thus partly function as taxes on labor, except that they only apply to the subset of workers covered by the $(L_{CSI} + L_{JS} + L_{MW})$ regime, so that they are really taxes on salaried or dependent labor. In parallel, many countries have special regimes merging L_{CSI} with T_{PIT}, or even with T_{CIT}, particularly for individuals around the fuzzy border between self-employment and the micro firm. These regimes are very important for social protection, and we discuss them further in chapter three.

In sum, $T(.) = T(T_{PIT},..., T_{ENF})$ is an abstract representation of the laws and institutions that set and collect taxes and make transfers to households. As with L(.), T(.) also contains important elements of countries' social protection policies, quite directly through transfer programs in T_{TRN}, but also indirectly through exemptions in T_{VAT} or—less noted but equally important—policies to promote micro and small firms through special regimes. Of course, T(.) also captures efforts to lessen income inequality via the progressivity of T_{PIT}.

Again, while there may be similarities, the "realm of taxes and transfers" in Latin America differs from that of most OECD countries. First, special regimes merging L_{CSI} with T_{PIT} or T_{CIT}, or setting up special rates for each tax, are very important in the region. Second, as discussed in chapter four, with the exceptions of Argentina and Brazil, tax to GDP ratios are lower because: (i) VATs generally have lower rates and more exemptions; (ii) revenues from personal income taxes are lower due to loopholes for high-income earners, and reduced withholding by firms of workers' taxes due to reduced coverage of L_{CSI}; (iii) property and environmental externalities are subject to lower taxes; (iv) energy is subject to lower taxes or, in many cases, is subsidized; and (v) institutions are less efficient at collecting taxes, resulting in more evasion.

2.1.3. The realm of market conditions

The third broad area making up the environment is related to the functioning of markets, which we label the "realm of market conditions," symbolized by $M(M_{FTI}, M_{COM}, M_{PDP}, M_{ENF})$. It is again very broad, but for our purposes we highlight four dimensions:

- M_{FTI}: laws regulating foreign trade and investment—including free trade agreements—determining the degree of competition from abroad faced by domestic firms, the prices paid by consumers for tradeable goods, and the opportunities that domestic firms have to export and participate in world value chains;
- M_{COM}: laws regulating domestic competition—including provisions to punish monopolistic behavior—that determine the extent of market contestability, particularly in areas not exposed to foreign competition;
- M_{PDP}: policies—sometimes labelled productive development policies and sometimes industrial policies—to promote specific sectors, activities or even individual firms, be they in the form of subsidies to innovation, subsidized credit from development banks, or measures to correct coordination, information or other types of market failures; and,
- M_{ENF}: the functioning of institutions enforcing credit or commercial contracts between parties, including bankruptcy procedures and the mechanisms to appropriate collateral.

In sum, $M(M_{FTI}, ..., M_{ENF}) = M(.)$ is an abstract representation of the laws and institutions that determine the functioning of markets, including the institutions designed to provide legal certainty to parties in an exchange, correct market failures or promote certain sectors or activities.

As with L(.) and T(.), each dimension of M(.) consists of many laws and institutions, but it is not necessary here to discuss them further. In this essay the various dimensions of M(.) will by and large be in the background, although in chapter three we will return to some of them in the context of discussions on firm behavior and related determinants of productivity.

Finally, and yet again, the "realm of market conditions" is very different in Latin America than in the OECD. For instance, fewer public resources are channeled to technology adoption and innovation. In parallel, institutions that enforce competition policy are weaker, as are institutions that enforce contracts.

2.1.4. Institutions and social and economic outcomes

The three realms mentioned above jointly form a country's environment, symbolized by:

$$(1)\ \mathbf{E}\{.\} = \{\ L(L_{CSI}, L_{JS},..., L_{ENF}),\ T(T_{PIT}, T_{CIT}, ..., T_{ENF}),\ M(M_{FTI}, ..., M_{ENF})\ \}$$

We highlight, first, that $\mathbf{E}\{.\}$ captures the functioning of institutions that enforce laws and thus determine whether the rule of law is strong or weak. Put differently, $\mathbf{E}\{.\}$ reflects what individuals consider to be the *de facto* rules of the game, which can differ from the *de jure* rules depending on how they perceive the efficiency of the institutions in charge of collecting taxes and social insurance contributions, and the fairness of the courts resolving labor, credit, commercial or competition disputes between workers and firms, between firms, between firms and banks or between firms and the government.

Second, $\mathbf{E}\{.\}$ reflects the functioning of the institutions in charge of service delivery (for health or childcare in particular) or regulating savings for pensions. These institutions can be public or private and some may be run by subnational governments. What matters for our purposes is the quality of the services provided or the efficacy of the regulations, which are important determinants of the value that individuals derive from participating in the (L_{CSI} + L_{JS} + L_{MW}) regime or from receiving benefits through non-contributory or transfer programs.

Institutions and policies that shape social outcomes overlap with those that determine the efficiency with which labor, credit and output markets operate. For example, institutions in (L_{CSI} + L_{JS} + L_{MW}) are central to workers' welfare as they determine their access to contributory social insurance and coverage of job stability provisions and minimum wages; but they are also critical determinants of firms' labor costs and hiring decisions. Corporate taxes, including special regimes, determine firms' profitability, but also the number and size of firms that make up the demand side of the labor market. Similarly, the institutions in charge of combating monopolies impact the distribution of income as they determine the rents that can be captured by abusive behavior, but they also impact the efficiency of product markets. And the regulations on international trade determine the degree of foreign competition faced by domestic firms, but also the prices domestic consumers pay for key staples like grains or meat.

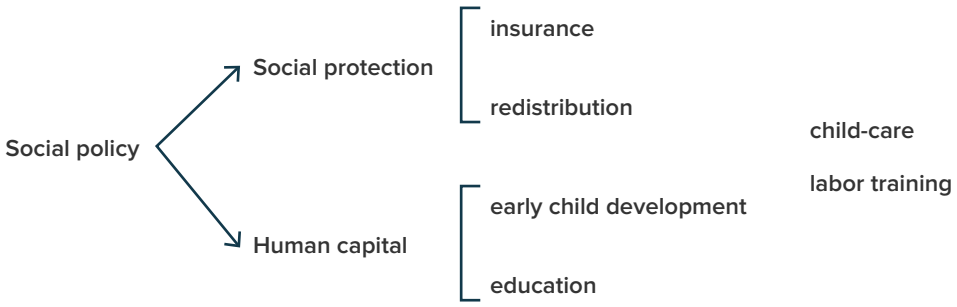
Countries are distinguished by their environments. Computers, warehouses, machine tools, trucks—capital goods—are pretty much the same in Colombia and Peru. Chilean and Argentinian electrical engineers know the same laws of physics, and Ecuadorian and Mexican workers with primary education have similar skills or at least could acquire them under the same conditions. But Colombia's $\mathbf{E}\{.\}$ is different from Peru's. They may have similar L(.)s, but different T(.)s. And even if their L(.)s are *de jure* the same, *de facto* they are not because their enforcement differs.

Institutions, laws and policies in $\mathbf{E}\{.\}$ are products of history. They reflect ideas and aspirations but also the balance of power between different groups in society; together, they shape everybody's *modus vivendi*, and everybody's welfare.

2.2. Social Protection Policies

We now introduce the concept of social protection and relate it to $\mathbf{E}\{.\}$. Graph 1 highlights the distinction between social policies made in the introduction.

Graph 1: Social Protection and Investments in Human Capital



Source: authors' elaboration.

Although policies to invest in human capital are as important as social protection policies, we do not consider them in this essay. We also ignore other policies that are indispensable for equity, like those aimed at preventing racial discrimination. However, sometimes policies cannot be easily classified under a single category. In Graph 1, for instance, childcare is relevant for early child development, but also for women's participation in the labor market. Similarly, labor training increases workers' human capital, but also impacts their earnings in the labor market. We also ignore those policies here.

In this essay we focus only on the social protection policies identified at the top of Graph 1. These policies have two main objectives (Ferreira & Robalino, 2011):

- to protect households against risks; and,
- to mitigate inequalities and redistribute to those in need.

The first objective is usually pursued through health, disability, unemployment and life insurance and retirement pensions—jointly known as social insurance policies— and the second through tax and transfer policies. That said, in practice it is impossible to make a unique association between objectives and policies because: (i) social insurance policies also impact income distribution; (ii) taxes and transfers affect social insurance coverage; and (iii) some risks in the region are sometimes covered by transfer rather than social insurance programs.

2.2.1. Identifying social protection policies

What laws and institutions in $\mathbf{E}\{.\}$ constitute social protection policies? Social insurance is usually associated with L_{CSI}. But in Latin America: (i) L_{CSI} occasionally provides more than insurance because some countries also finance with earmarked wage-based taxes programs unrelated to risks, like childcare, housing, education, workers' training or child allowances; and (ii) social insurance is also provided by programs in L_{NCSI}.

Importantly, the risks of loss of employment are usually not covered through a program in L_{CSI}; as noted, only Argentina and Chile have unemployment insurance funded from wage-

based contributions. (Brazil also has unemployment insurance, but it is funded by other taxes.) Rather, these risks are covered through regulations on job stability in L_{JS}. These regulations are often considered part of labor regulations and not part of contributory social insurance because there is no flow payment out of wage-based taxes specifically earmarked to compensate workers when they are separated from their jobs. But the absence of an explicit flow payment does not mean that these regulations do not affect firms' expected labor costs and provide some benefits to workers in return, partly mimicking the role of insurance, except that payments are made *ex post* and not *ex ante*.

In parallel, redistribution is usually associated with personal income taxes and transfer programs in T_{TRN}. But exemptions to the VAT, very relevant in some countries in the region, are motivated by redistributive objectives. Less noticed—but probably more important in some countries—are special regimes for micro and small firms, generally associated with supporting low-income entrepreneurs.

Social protection policies vary from country to country, precluding a definition that fits all. Nevertheless, it is useful to identify the main ones and match them with their main objectives, in the understanding that this match does not apply strictly to any country. We thus have:

Social insurance policies: L_{CSI}, L_{JS} and L_{NCSI}

(2) **Social protection policies:**

Redistribution policies: T_{PIT}, L_{MW}, T_{TRN}, and, in part, T_{CIT} and T_{VAT}

Relation (2) helps to make two observations on social protection policies:

- they are reflected in programs that insure against risks and in programs that provide benefits independently of risks; and,
- they operate partly through the labor market and partly through the tax and transfer system.

Transfer programs in T_{TRN} can be universal or focused on a subset of individuals. In the latter case, when the targeting criteria involves incomes below some threshold, they are called social assistance or poverty programs (the most common of which are conditional cash transfer programs or CCTs). On the other hand, transfers can be in-kind or in cash. In-kind transfers may be the distribution of some key staples like milk, or services like childcare, although in some countries they also include health services (which is subsequently not included as an insurance program in L_{CSI} or L_{NCSI}).

In Latin America there is large variation in the programs and benefits included in L_{CSI}, L_{JS}, L_{NCSI} and T_{TRN} and the regulations to qualify for each. There is also variation in the parameters of L_{MW}, T_{PIT}, T_{CIT} and T_{VAT}. Consider, for instance, health services. Brazil provides all households with services of the same quality through a transfer program in T_{TRN}; Ecuador and Peru provide these services through a combination of social insurance programs in L_{CSI} and L_{NCSI}, with differences in quality between them. This implies that in Brazil there are no wage-based earmarked taxes for health services, as they are all financed from general revenues, unlike Ecuador and Peru, where they are financed by a combination of these taxes (when L_{CSI}) and other sources of revenues (when L_{NCSI}). It also implies that access to health services is not a

relevant consideration when firms and workers engage with each other in Brazil, as opposed to Ecuador and Peru.

Now consider now regulations on job stability, L_{JS}. Colombia, Ecuador, Honduras, Mexico, Peru and many other countries do not allow firms to dismiss workers when demand falls or labor-saving technical change occurs. These are not considered just causes of dismissal, and workers can take firms to court to be reinstated in their jobs or receive compensations. In Chile, however, changes in demand and technical change are accepted as just causes of dismissal.

Take retirement pensions. Through a universal program in T_{TRN}, Mexico offers a pension to all its elderly, and through a targeted program in T_{TRN} Colombia restricts these pensions to low-income elderly without contributory pensions. This implies that in Colombia, but not in Mexico, workers' earnings and status in the labor market prior to retirement have substantive implications for the coverage of these pensions.

Next, consider income support for the poor. In Ecuador this is done through a targeted transfer program in T_{TRN}, conditional upon workers not benefiting from L_{CSI} programs; in Brazil this is also done through a similar targeted program in T_{TRN}, but independent of workers' participation in L_{CSI} programs. Or, lastly, consider consumption taxes in T_{VAT}. Mexico' s VAT has preferential rates for food, medicines and other goods accounting for close to 50 percent of the consumption basket; in Chile VAT rates are the same for practically all goods.

The distinction between contributory and non-contributory programs also varies across countries. In some, like Chile, resources contributed by workers to their individual retirement accounts are complemented at the time of retirement with government contributions to increase the size of the pension; one could thus label them semi-contributory retirement pensions. A similar situation occurred in Argentina when subsidies were introduced to allow workers to obtain a contributory pension even though they did not qualify for one. In Colombia and Mexico, contributory health programs receive resources from the government, in addition to contributions from firms and workers.

More examples could be listed, and from this perspective, it could be better to think of a continuum of programs. That said, in this essay we use the contributory/non-contributory classification with the understanding that it is not always a sharp dichotomy and specific circumstances need to be considered in each case. Our classification is based on two defining features of contributory programs: (i) they are mostly funded from obligatory contributions by workers or firms that are proportional to workers' earnings; and (ii) benefits take the form of legal entitlements. On the other hand, non-contributory programs are mostly funded from government resources, participation is voluntary, and benefits may—but do not have to—be legal entitlements.

As mentioned, transfer programs in T_{TRN} can target a subset of households or be universal. In the first case, the targeting criteria is usually income (or some proxy); in the second, all households are included. Targeted transfers in T_{TRN} are distinguished from L_{NCSI} programs because the latter focus on all workers not covered by L_{CSI} programs, regardless of income level. The distinction matters because they respond to two different circumstances: lack of coverage of L_{CSI} programs for workers that are not necessarily poor; and poor households, even if some of their members have jobs covered by L_{CSI} programs. Of course, there is an important overlap between households lacking coverage of L_{CSI} programs and households that are poor, contributing to the conflation of non-contributory programs and targeted transfers. But this overlap is not perfect because there are some poor households covered by L_{CSI} and some non-poor households that benefit from L_{NCSI} programs (in fact, in most cases, more than poor households).

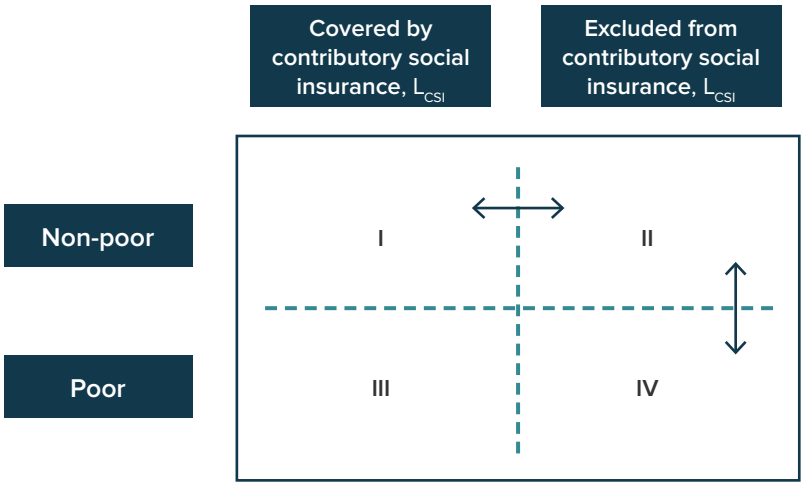
That said, practices vary and regulations are often imprecise. Honduras has a health program in T_{TRN} for its vulnerable population, a concept never defined in its laws, so that in practice all informal workers have access, making it *de facto* an L_{NCSI} program. In Colombia, only poor informal workers can benefit from its non-contributory health program, thus making it doubly targeted on income level and labor status; the program can be thought of as a non-contributory program in L_{NCSI} targeted to the poor or as a transfer program in T_{TRN} focused on workers that are both poor and informal. In Mexico, all workers have access to a non-contributory pension, so it is a universal transfer in T_{TRN}, while only those without access to social security (i.e., coverage of L_{CSI}), regardless of income level, can benefit from its non-contributory health program, making it a program in L_{NCSI} (like in Ecuador and Peru). Argentina's non-contributory health program has the same characteristics as Mexico's, a program in L_{NCSI}, with access conditioned on labor status, not income, and services delivered by the public health system (often, as in Mexico, with quantity constraints).

These examples are sufficient to highlight three important dimensions in which countries' social protection policies differ:

- the balance of programs between those that operate through the labor market and those that operate through the tax and transfer system;
- the balance between universal and targeted transfer programs; and,
- the combination of labor market status and income level as targeting mechanisms to qualify for benefits.

Graph 2 maps workers into four quadrants. The columns separate them by their access to L_{CSI} programs; the rows do so by income levels. Following the discussion above, coverage of individual social protection programs depends on where workers are located.

Graph 2: Income Levels, Labor Status and Access to Social Protection



Source: Levy (2008).

Universal transfer programs benefit workers in all quadrants, like health services in Brazil. On the other hand, transfer programs targeted by income levels benefit only workers in quadrants III and IV, like Mexico's CCT. However, when the rules of the CCT exclude workers covered by L_{CSI}

programs, as in Ecuador and Peru, only workers in quadrant IV receive them. As a result, transfer programs can be universal, targeted by income level only, or by income level and labor status.

Thus, while there are broad similarities in social protection policies across Latin American countries—in particular, the provision of insurance through a mix of contributory and non-contributory programs and the reliance on targeted transfer programs for the poor—many important features are also country specific. This, together with the fact that policies are implemented in the context of each country's $\mathbf{E}\{.\}$, helps to explain why social protection outcomes differ in the region.

Finally, and critically for the purposes of this essay, the location of the dotted lines on Graph 2 is not exogenous; it depends on social protection policies themselves, which, as illustrated by the two-way arrows, can move them in either direction; and, of course, on all other elements of $\mathbf{E}\{.\}$. We return to this point below.

2.2.2. Formality and informality: workers

The concepts of formality and informality loom large in discussions of social protection in Latin America, and it is useful to relate them to Graph 2. Throughout the essay we use the following definitions:

- formal workers are those covered by contributory social insurance L_{CSI}, in quadrants I and III in Graph 2; and,
- informal workers are all others, in quadrants II and IV.

In general, informal workers are covered by non-contributory social insurance programs, although the significance of this coverage varies across countries. However, when formal workers are in a relation of dependency and subordination with a firm, they are also subject to regulations on job stability and minimum wages. This implies that—at least on paper—formal workers, particularly dependent ones, have much broader protections against risks than informal workers.

We stress that formality is defined with respect to coverage of L_{CSI}, not compliance with tax laws or some other requirement (like having a bank account). Unfortunately, the terms formal and informal are sometimes used depending on whether workers pay taxes or are registered in some administrative database, causing confusion. More generally, usage suffers from conceptual stretching: the same words describe different phenomena. As Guha-Khasnobis et al. (2006) write, the words formal and informal "are better thought of as a metaphor that conjures up a mental picture of whatever the user has in mind at a particular time."

Of course, the words formal and informal may be defined differently, and for other purposes this may be correct. However, from the point of view of social protection, what matters is the extent to which individuals are insured against risks, not whether they pay taxes, comply with this or that registration requirement, or have a bank account. Thus, for the purposes of this essay, we use these words in the sense spelled out above.[3]

[3] This is more than academic precision. Many workers in the region are informal but pay taxes. A high earning self-employed doctor may pay their income taxes but is informal if they do not contribute to L_{CSI} (as is usually the case).

From a legal point of view, there are three different situations that cause workers to be informal. The first occurs when a country's (L_{CSI} + L_{JS} + L_{MW}) laws do not require the self-employed to contribute (although this sometimes depends on income levels); the second happens when firms hire workers under a relation of dependency and subordination but break the law. In the first situation, workers are informal and legal; in the second, they are informal and illegal, although the firm is responsible for the illegal act. The third situation arises when the requirement to contribute is unclear due to ambiguities in (L_{CSI} + L_{JS} + L_{MW}) laws. In this situation, the legal status of informal workers is equally unclear. Thus, there is heterogeneity in the legal status of informal workers in quadrants II and IV of Graph 2.

2.2.3. Formality and informality: firms

The concepts of formality and informality are unambiguous when applied to workers. At any point in time, an individual is either contributing to L_{CSI} or not (although they may under-report earnings to minimize contributions). But they are ambiguous when applied to firms. How does one classify a firm with 10 workers, where four are covered by L_{CSI} and six are not? Clearly, there is a continuum, with a somewhat arbitrary border separating formal from informal firms. In this essay, we chose the extreme of this continuum and use the following definitions:

- informal firms are those with no workers covered by L_{CSI}; and,
- formal firms are those with at least some workers covered by L_{CSI}.

As with workers, informal firms may be legal, illegal or somewhere in the middle. In countries where the self-employed *de facto* run a micro firm but are not required to contribute to L_{CSI}, the firm is informal and legal. In countries where only dependent workers are required to contribute, firms engaging with non-dependent workers, like most family firms, are also informal and legal. On the other hand, if firms are required to comply with (L_{CSI} + L_{JS} + L_{MW}) laws but fail to do so, they are informal and illegal. However, firm behavior can be more complex, mixing workers with and without relations of dependency and subordination, or sometimes mixing short- and long-term contracts to elude (and not explicitly evade) these laws; they can also sub-contract some tasks to another firm or evade along other dimensions.[4]

2.2.4. Formality and informality: legality and productivity

We make three points regarding the previous definitions. First, informality is not equivalent to illegality, unless a country's laws require everybody to contribute (as in Argentina and Brazil).

Second, because (L_{CSI} + L_{JS} + L_{MW}) laws in Latin America are not enforced effectively, or because they are ambiguous, *de facto* some individuals behave as if they are not subject to them. When this behavior is generalized, it can be accepted and tolerated by all, including the authorities

[4] The extent of informal firms' illegal behavior varies across countries and occurs through various margins: not paying workers' contributions to L_{CSI} or paying, but under-declaring wages. For example, see Ulyssea (2018) for a discussion of Brazil, where all firm informality is illegal; Levy (2018) for a discussion of Mexico, where firm informality is mostly legal; and Bergolo and Cruces (2014) for a discussion of Uruguay, where firms partially break the law.

in charge of enforcement.[5] Generalized non-compliance among some firms and workers then becomes a tacitly accepted social norm. Few expect a two-person firm in Honduras, Ecuador, or any country for that matter, to comply with ($L_{CSI} + L_{JS} + L_{MW}$) laws, even if there is a clear relation of dependency and subordination; but everybody expects a fifty-person firm to do so. Similarly, when all the self-employed are required to contribute, few expect low-income individuals to comply, but all expect high-income ones to do so. From this perspective, informality is the sometimes legal, sometimes illegal and other times ambiguous status of firms and workers that results from a context where laws may be imprecise, their enforcement is weak and tacitly accepted social norms sanction behavior that could arguably be illegal.

The third point is central for this essay. Firms' formality status matters for two related but conceptually different reasons: social protection and productivity. It matters for social protection because it speaks to the social protection benefits that their workers receive. What is relevant here is what happens to workers, not the firm itself, since firms do not benefit from health services or pension payments. Firms' formality status also matters for productivity because it speaks to the relative prices and incentives that they face when making critical business decisions like which technologies to adopt, how many workers to hire and under what contractual modalities, and so on. In other words, what is relevant here is firm behavior along the dimensions that impact productivity, regardless of workers' access to social benefits.

These two reasons are at times conflated, leading to the implicit assumption that if workers in an informal firm are induced to participate in contributory programs, then the behavior of the firm will change along various dimensions that impact productivity. In other words, if informal workers formalize, then the informal firm automatically faces the same incentives as a formal one. But this depends on the mechanisms deployed to induce workers to formalize. If they ensure that the informal firm faces the same incentives and relative prices as formal ones, the assumption is correct. But, as we discuss in the next chapter, this is not always the case. More often than is at times recognized, countries in Latin America deploy mechanisms to formalize informal workers that imply no changes in the underlying incentives and prices faced by informal firms (even if now they are labelled formal). These mechanisms, sometimes called formalization programs, are supported by a myriad of special tax and contributory regimes that for all intents and purposes leave informal firms facing the same incentives and relative prices. In these cases, workers' access to social protection will improve, but without any implications for firm behavior and productivity.

To sum up: because the words formal and informal are deeply engrained in social protection discussions in Latin America, it is essential to be precise in their usage. In this essay, these words are defined with respect to workers' access to social insurance. They should not be equated with legality and illegality. These words are also applied to firms. This usage is fine, as long as there is clarity that the words do not necessarily imply anything about their productivity. What matters for productivity are the relative prices and incentives faced by firms, which is often—but not always—correlated with their formality status.

[5] The behavior of enforcement agencies may be rational. In the case of small firms, collection costs may exceed the amounts collected. Agencies may also be criticized for going after "small guys" rather than "bigger fish."

2.2.5. Labor market outcomes and social protection

For a simple but powerful reason, the formal-informal dichotomy is central for social protection in the region: insurance coverage depends on the formal or informal status of workers; and, in some countries, also access to transfer programs.

Indeed, if all social protection policies operated through T(.) and none through L(.), workers' formal or informal status would be irrelevant from the perspective of social protection (indeed, the words formal and informal, as defined here, would be meaningless). Alternatively, if all transfers were independent of workers' labor status, and if the coverage against risks offered by programs in L_{NCSI} were the same as those in ($L_{CSI} + L_{JS} + L_{MW}$), workers' labor status would also be immaterial for social protection since all workers would receive the same transfers and be equally protected against risks.

But this is not the case in Latin America. By and large, a substantial share of social protection coverage occurs through programs in L(.) relative to programs in T(.). And, within L(.), there are differences in the quality and generosity of benefits between ($L_{CSI} + L_{JS} + L_{MW}$), on the one hand, and L_{NCSI}, on the other. Further, some transfers in T(.) are conditioned on formal and informal status. As a result, in Latin America, labor market outcomes are critical determinants of social protection outcomes.[6]

What determines labor market outcomes? The number and human capital of individuals participating in the market on the supply side, and the number and type of firms and the modalities of their engagements with workers on the demand side. The impact of human capital on the supply side has received substantial attention, and we do not discuss it here. Rather, we focus on the role of firms on the demand side.

2.2.6. Firms and social protection

Graph 3 highlights the interdependence between firms and social protection: social protection outcomes depend on firms' behavior, and firms' behavior depend on social protection policies. The graph lists some elements of this interdependence; a discussion of the mechanisms is provided in chapter three. The importance of this point for social protection cannot be overestimated: the demand side of the labor market depends on social protection policies.

Of course, many policies affect firm behavior, not only social protection ones: regulations on international trade, domestic competition, access to credit, innovation and so on. Put differently, firm behavior depends on policies in the realm of market conditions M(.), in addition to those in the realms of entrepreneur-worker relations L(.) and taxes and transfers T(.) or, in short, on the environment $\mathbf{E}\{.\}$. Graph 3 focuses on social protection policies because of the interests of this essay, but evidently all other policies matter.

In parallel, firm behavior determines many outcomes, not only social protection ones: the type of goods exported, the pace of technology adoption and innovation, and so on. All these outcomes

[6] Public primary education provides a useful contrast because it is a universal transfer program in Latin America. Outcomes in the labor market have no bearing on coverage because access is the same for all regardless of workers' labor status.

in turn impact the productivity of the economy and the rate of GDP growth. But the point we want to emphasize here is that if social protection policies facilitate firm performance, they contribute to good economic outcomes, which in turn lead to good social protection outcomes; but if they obstruct it, they hurt economic outcomes and, in turn, social protection outcomes.

Because firms play a large role in determining social protection and economic outcomes, it is indispensable to pay close attention to the links between them and social protection policies. Unfortunately, analyses of social protection in Latin America sometimes overlook this point, or at least fail to give it the importance it deserves. But the point is central.

Social protection outcomes in Latin America and the OECD differ for many reasons, but a key one is associated with differences in the number, type and behavior of firms in each region. As elaborated below, the majority of firms in Latin America are informal and have less than five workers. They require few, if any, highly skilled workers; many have short lives, and those that survive rarely grow. Differences between the region and the OECD are accentuated when self-employed individuals are counted as firms (as they should be) because self-employment is much larger in Latin America and because the border between self-employment and a micro firm is very fluid. As we expand on below, widespread self-employment and a proliferation of micro firms are two reasons why social protection outcomes in the region are unsatisfactory, and also why productivity is low.

Graph 3: Interdependence between Firms and Social Protection

- Are there many firms without dependent workers?
- Do firms comply with (L_{CSI} + L_{JS} + L_{MW}) laws?
- Are there many micro or small firms where the relation of dependency is ambiguous?
- What is the skill composition of firms' labor demand?
- Do firms offer workers short or long-term contracts? How many firms invest in labor training?
- Do patterns of firm entry, exit and growth increase workers' productivity?
- How costly are contributions to L_{CSI}? What is the quality of benefits and the rules to access them?
- Do job stability regulations allow firms to adjust to negative shocks? Are minimum wages high or low?
- Is enforcement of (L_{CSI} + L_{JS} + L_{MW}) laws dependent on firm size? Can firms replace them with tacitly accepted social norms if they are small?
- Are there special tax regimes for low-income entrepreneurs? Do they vary with firm size?
- Do firms evade more when L_{NCSI} programs are present?
- Do one-person firms have to participate in L_{CSI} programs?

Source: authors' elaboration.

2.3 Social Protection Outcomes and Productivity Outcomes: A Unified Process

We now focus on the determinants of social protection and productivity outcomes.[7] We keep the discussion at a high level of generality. Assume there are n individuals in the country, indexed I_1, I_2, ..., I_n. Each has a level of human capital H_1, H_2, ..., H_n, so that the stock of human capital is I.H. The stock of physical capital is K so that:

(3) Available resources for production = (I.H, K)

In any year, resources are given, but change over time because of population growth, improvements in the education and abilities of individuals, or investments in physical capital.

Because we want to understand labor market outcomes, we are interested in the division of individuals between those that work and those that do not; and among those that do, between entrepreneurs, the self-employed and workers in firms. We are also interested in the number and type of firms present in the market, the technologies they use and the type of contracts they establish with their workers.

2.3.1. Firms and aggregate productivity

Let $\mathbf{T}$ denote the set of all technologies available, from the simplest to the most sophisticated and complex—a catalogue of all known methods to produce goods and services. Given $\mathbf{T}$, the specific technologies actually used depend on which firms exist and who manages them.

In this vein, let there be n firms with productivity levels TFP_i, i = 1, 2,.., n, depending on the technology chosen by each from all the available options in $\mathbf{T}$. In turn, let a_i be the share of the available resources—capital K and labor I.H in relation (3) —captured by the i^{th} firm. We can then write aggregate total factor productivity, **TFP**, as:

$$(4)\ \mathbf{TFP} = a_1TFP_1 + a_2TFP_2 + ... + a_nTFP_n\ ;\ \ \Sigma a_i = 1$$

or, stated in words, **TFP** is the average of the productivity of all firms, where the weights are the share of total resources captured by each.[8]

Relation (4) is just an accounting identity; it holds in any country and provides no explanation of behavior. But it helps to make three observations on the impact of the size and productivity of individual firms on **TFP**:

[7] Various models focus on the relation between informality and productivity; see, for instance, Galiani and Weinschelbaum (2011), Levy (2008, 2018) and Ulyssea (2020b). The discussion here should be thought as a framework that can encompass aspects emphasized by each.

[8] To simplify the presentation, we ignore the distinction between physical and revenue productivity. Here, TFP_i is the value added produced by firm i with one peso of capital and labor, reflecting the interaction between the physical efficiency of the production process and the price received by the firm for its product; see Hsieh and Klenow (2009).

- if all firms had the same TFP_i, their size —that is, the value of the a_i—would not matter; whether more or less resources were allocated to one or another would be immaterial for **TFP**. But if differences in firms' TFP_i are large, then the allocation of resources between them matters a lot for **TFP**;
- if all firms had the same TFP_i, their label—that is, whether they are formal or informal—would be immaterial for **TFP**. But if there are systematic differences in their TFP_is, the allocation of resources between them matters a lot for **TFP**; and
- from the point of view of **TFP**, a single micro or small firm is almost irrelevant, regardless of its TFP_i. However, if there are many of them, their sum will matter**.**

What do we know about relation (4) in Latin America? Unfortunately, the data to characterize it fully are insufficient, because no country has an economic census covering firms of all sizes in all areas of economic activity. That said, we can combine data from employment and firm surveys to provide an incomplete but nonetheless useful characterization. We do so from three perspectives: data on the composition of employment, data on firm size and data on firm productivity.

2.3.2. The composition of employment

Employment surveys in the region usually ask workers whether they work in a firm or are self-employed and, in the first case, the size of the firm that they work for. Table 1 presents this data for the countries considered in this essay and, in the last row, for the United States.

Table 1: Distribution of Employment
(shares)

	Self-employment	Employment in firms
Argentina	27%	24% in up to five workers 18% in 100 or more
Brazil	34%	15% in up to five workers 28% in 50 or more
Chile	21%	20% in up to nine workers 33% in 50 or more
Colombia	34%	25% in up to five workers 26% in 50 or more
Dominican Republic	44%	14% in up to 10 workers 12% in 50 or more
Ecuador	35%	41% in up to 10 workers 10% in 100 or more
Honduras	54%	22% in up to 10 workers 15% in 150 workers or more
Mexico	21%	30% in up to five workers 21% in 50 or more
Peru	37%	35% in up to 10 workers 9% in 100 or more
United States	10%	6% in up to five workers 56% in 100 or more

Source: for the United States, Statistics of United States Business, Census Bureau.

Despite the fact that the data are not fully comparable because definitions vary somewhat across countries, as do the reported ranges of firm size, three facts about the region stand out:

- self-employment is very widespread, sometimes accounting for over half of the labor force, but in all cases, it is at least double that of the United States;
- a large share of workers is employed in firms with up to five (or 10) workers; and,
- the share of workers in firms with 50 or more employees is substantially lower than in the United States.[9]

The bottom line is that employment in the region is extremely dispersed. Millions of individuals work on their own and millions more in small firms. These facts are insufficiently captured by surveys or censuses of firms, either because they ignore self-employment, or because they do not capture many micro or small firms that carry out their activities in the streets or in the household.

That said, it is useful to recall that the border between self-employed workers and micro firms is fuzzy because many "self-employed" are aided by other workers. Thus, in Table 1 it is best to add the share of the labor force that is self-employed to the share of those employed in firms with up to five (or 10) workers. With the exception of Chile, in all countries at least half of the labor force is in the self-employed cum micro firm category.

2.3.3. The size distribution of firms

Data on firms in the region are scarcer than data on employment. Firm censuses sometimes only include those with 10 or more workers, and sometimes only cover manufacturing. But it is possible to piece together a reasonable picture for a few countries. Table 2 provides one for Ecuador, Mexico, Peru and for comparison again, the United States. These data exclude self-employment.

The contrast between Latin American countries and the United States is sharp. Take, for instance, Ecuador. In 2019, its population was one-twentieth of the US population and its GDP one-two hundredth, but the number of firms was 38 percent of the United States. In both countries, firms with up to 10 workers (nine in the US) made up the majority, but accounted for 99 percent in one and 60 percent in the other. In Ecuador, those firms accounted for 79 percent of workers, while in the United States they account for only 10 percent. The comparison with Peru is almost as dramatic: its population was one-tenth and its GDP one-hundredth of the United States, but it has 40 percent of the number of firms. These facts

[9] CAF (2018) compares the composition of salaried employment between the United States and the Latin American average and finds that in the United States, 12 percent of employment occurs in firms with up to 10 workers, compared to 48 percent in the Latin America. IDB (2010) presents the following comparison only for manufacturing: in the United States, 54 percent of all firms have nine or fewer employees, and account for only 4.2 percent of employment, while 9 percent have 100 or more employees and account for 68 percent of employment. In Argentina, 84 percent of all firms have nine or fewer employees, accounting for 22 percent of employment; in Bolivia, 92 percent of firms and 44 percent of employment; and in Mexico 90 percent of firms and 23 percent of employment. At the other end, in Argentina only 0.2 percent of firms have more than 100 employees, while the numbers are 0.6 percent in Bolivia and 1.6 percent in Mexico. These comparisons exclude self-employment. The World Bank (2014) notes that Latin America differs from other regions of the world in that, given its income per capita, a much larger share of the labor force is self-employed.

complement those presented in Table 1 showing that self-employment is substantially higher in Peru (37 percent) and Ecuador (38 percent) than the United States (10 percent).

Unfortunately, the data for Ecuador and Peru do not allow us to examine firms with more granularity. But the data from Mexico do: of the 4.5 million firms with 10 or fewer workers, 4.2 million only had five or fewer. It stands to reason that Ecuador and Peru are similar, a supposition buttressed by the data on the composition of employment shown in Table 1.

Considered jointly, Tables 1 and 2 yield two very important conclusions about the region:

- in addition to widespread self-employment, there is a huge number of firms with tiny a_is, that is, micro and small firms; and,
- the sum of the a_is of these firms absorb a large share of countries' resources.

Table 2: Distribution of Firms
(In thousands)

	Firm size by number of workers	Number of firms	Share	Number of workers	Share
Ecuador	2–10	2,987	99.6	5,764	79.3
	11–100	11	0.03	727	10.0
	101 +	0.1	0.01	773	10.0
	Total	**2,998**	**100.0**	**7,264**	**100.0**
Mexico	2–10	4,535	95.0	9,596	43.5
	11–100	211	4.40	4,383	19.9
	101 +	27	0.60	8,044	36.5
	Total	**4,773**	**100.0**	**22,024**	**100.0**
Peru	2–10	3,163	98.9	8,586	73.0
	11–100	29	0.08	1,563	13.3
	101 +	4.1	0.02	1,610	13.7
	Total	**3,196**	**100.0**	**11,759**	**100.0**
United States	2–9	4,726	60.1	12,503	9.7
	10–99	1,405	17.9	29,851	23.2
	100 +	1,729	21.9	86,147	67.0
	Total	**7,861**	**100.0**	**128,592**	**100.0**

2.3.4. The productivity of firms

To characterize relation (4) it is also necessary to provide data on the productivity of individual firms, that is, the values of TFP_i. These data are difficult to obtain because complete censuses including capital stocks and other variables needed to compute TFP_i are rare in Latin America. Some studies have done so, but often only for firms that have 10 or more workers (which, as seen, leaves out a substantial share of resources and firms), or only in manufacturing.

Nevertheless, there are various regional and country studies that shed useful light. Early ones are IDB (2010), Busso et al. (2013) and World Bank (2014), which all find large productivity differences between formal and informal firms. More recently, CAF (2018) finds that, in general, formal firms are about 45 percent more productive than informal ones. Ulyssea (2018) also shows large productivity differences between formal and informal firms in Brazil. Levy (2018) finds that in Mexico, after controlling for size and location, formal firms are between 35 and

50 percent more productive than informal ones, even comparing firms within the same six-digit sector classification. All in all, these studies converge on two key findings:

- the majority of firms with low TFP_i are micro or small; and,
- most micro and small firms are informal.

The bottom line is this: countries' $\mathbf{E}\{.\}$s allow many firms with low TFP_i to exist in the market and take away market share and resources from firms with higher TFP_i; and the common denominator among these firms is that they are mostly small and engage with their workers outside of the domain of the $(L_{CSI} + L_{JS} + L_{MW})$ regime.

Two additional observations can be derived from these studies. First, informal firms produce goods or services that are similar to those produced by formal ones. In the case of Mexico, for instance, Levy (2018) computes the number of six-digit sectors where informal firms account for more than half of all firms and finds that this occurs in 142 out of 279 sectors in manufacturing, 125 out of 154 in commerce, and 227 out of 258 in services; in other words, informal firms are pervasive across activities.

These results should not be surprising. Land transportation services can be produced by a one-person firm with one truck and one driver, or by a firm with, say, 100 trucks and 100 drivers. The services are pretty much the same—moving people or goods from A to B—but the technologies used can be very different, for instance, using GPS and related tools to optimize routes. Arepas and tortillas can be produced by industrial processes or by hand; clothing can be produced in the garage of a house with old sewing machines, or in a factory using automated machines; bricks can be baked in small ovens fueled by burning tires or in large ovens fueled by burning gas, and so on. The key point is that productivity differences reflect the fact that formal and informal firms have different TFP_is, not necessarily that they are in different sectors of the economy.[10]

The second observation is that these studies leave out self-employment. However, from the point of view of productivity, self-employed workers should be thought of as one-person firms. The fact that they are not captured in censuses or surveys of firms is immaterial. For the purposes of relation (4), they are like a firm, using part of a country's resources to produce output with a given productivity; in other words, they have their corresponding a_i and TFP_i. Since most self-employed workers are informal, and since self-employment is widespread, it is likely that the aforementioned studies underestimate the **TFP** losses associated with the region's informal sectors.

Summing up, if we compared relation (4) between a "typical" country in Latin America and a "typical" country in the OECD, we would find that in the former:

- the share of resources in one-person firms, or in very small firms with two to five workers, is proportionately much larger;
- conversely, the share of resources captured by firms with, say, 50 or more workers, is proportionately much smaller;

[10] There are differences in the TFP_is of firms producing similar goods everywhere, but they are much larger in Latin America. For instance, comparing firms in manufacturing at the four-digit level of disaggregation between the United States and Mexico, one finds that in the United States the difference in TFP_i between firms in the 10th and 90th percentile of the productivity distribution is 92 percent versus 173 percent in Mexico (Levy, 2018 and Hsieh & Klenow, 2009).

- differences in TFP_i between firms producing the same or similar goods are much larger; and,
- most low-productivity firms are outside the domain of (L_{CSI} + L_{JS} + L_{MW}) laws.

In parallel, focusing on social protection, we would find that compared to the "typical" OECD country, in the "typical" Latin American country:

- the vast majority of entrepreneur-worker-owners in one-person firms and workers in small two-to-five person firms are not covered by contributory social insurance programs because (L_{CSI} + L_{JS} + L_{MW}) laws exclude them, are ambiguous, or are evaded; and,
- workers protected by (L_{CSI} + L_{JS} + L_{MW}) laws are concentrated in medium and larger firms.

This discussion is central for social protection in the region. If there was less self-employment, and if there were fewer but larger firms, social protection outcomes would improve. At the same time, if many low-productivity firms did not exist, and the workers in those firms were absorbed by higher-productivity firms, **TFP** would be higher. In Latin America, low **TFP** and poor social protection outcomes go hand-in-hand.

2.3.5. Simultaneous determination of social protection and productivity outcomes

Relation (4) is the result of the process where an economy's resources (I.H, K) are combined to form firms (including one-person firms), who in turn deploy specific technologies from $\mathbf{T}$. It is useful to denote this allocation as **R**, and think of it as a listing of:

- individuals who do and do not participate in economic activity;
- the division of those participating between entrepreneurs, workers in firms, the self-employed and unemployed;
- the number of firms present, and the technology used by each;
- the number of workers hired by each firm and the type of contracts offered to them;
- the allocation of capital to each firm; and,
- the degree of compliance by firms and workers with various laws in L(.), T(.) and M(.).

Clearly, a given allocation **R** is associated with a corresponding **TFP**. Put differently, once we know **R**, we also know **TFP**, because we know who is doing what, how many firms there are, who manages them, what technologies are used and how much capital and how many workers of each skill level are in each. In fact, once we know this information, **TFP** is obtained by filling the corresponding numerical values in relation (4).

The process that allocates resources and determines an economy's productivity also determines individuals' income levels and access to social protection. Clearly, once we know **R** and **TFP**, we know the income **Y** = (Y_1, Y_2, ..., Y_n) of each individual, given who ended up being a worker, an entrepreneur or self-employed. We also know the compliance by firms and workers with (L_{CSI} + L_{JS} + L_{MW}) laws, and therefore their formal or informal status. With this information, we can map workers into the four quadrants of Graph 2 and determine the social protection programs **SP** = (SP_1, SP_2, ..., SP_n) to which each individual has access.

Of course, the process of "resource allocation cum productivity-determination cum income formation cum coverage of social protection" does not occur in a vacuum; it is shaped by a country's laws, institutions and policies, i.e., by $\mathbf{E}\{.\}$. Given the centrality of this process, it is useful to represent it through the following relation:

(5) [(l.H, K); $\mathbf{T}$; $\mathbf{E}\{$ L(.), T(.), M(.) $\}$] simultaneously determine [**R**, **TFP** and **SP**, **Y**]

In other words, given all the technologies available, a country's environment determines how individuals of given levels of abilities and education are distributed across occupations; how together with the available physical capital they form firms of different sizes using specific technologies; and, in turn, the aggregate productivity of the economy and individuals' incomes and social protection coverage.

We note outright that relation (5) excludes many factors that also influence resource allocation, productivity, income formation and social protection coverage. No consideration is given to macroeconomic policies or conditions in international financial or commodity markets, which play a critical role in Latin America. In the next chapter, in the context of the discussion on growth, we will extend relation (5) to consider them. At this point there is no need to do so since our focus here is on the structural determinants of social and economic outcomes, associated with a country's laws, institutions and policies.

Relation (5) is best interpreted as an abstract representation of the process by which a country's laws and institutions interact with its endowments of human and physical capital to determine how efficiently economic activity is carried out, how the fruits of that activity are distributed between those participating in it, and who is covered by its social protection programs.

Importantly, relation (5) should not be interpreted as an aggregate production function, where given $\mathbf{T}$, human and physical capital combine to produce output with maximum efficiency. On the contrary, the emphasis here is on the fact that the same human and physical capital (l.H, K) can produce different output levels—and thus result in different **TFP**s—depending on the rules of the game in $\mathbf{E}\{.\}$. In fact, two countries with identical resources and access to the same technologies, but different $\mathbf{E}\{.\}$s, would differ in their distribution of individuals across occupations, number and size of firms, technologies used and aggregate productivity, and, therefore, income distribution and coverage of social protection.

Of course, relation (5) is very complex and far from fully understood. Chapter three provides more structure on some aspects of this relation and sheds some light on how it works based on the empirical evidence from studies of different Latin American countries. Our objective at this point, however, is to highlight two points:

- the unity of the process that determines social protection and productivity outcomes; and,
- the critical dependence of that unified process on $\mathbf{E}\{.\}$.

What most distinguishes Latin America from other regions of the world is not $\mathbf{T}$, nor l.H and K; it is $\mathbf{E}\{.\}$. It has been amply documented that differences in income per capita between our region and other more prosperous regions are mostly due to differences in **TFP**, not

differences in factor accumulation.[11] These differences in **TFP** derive from differences in the region's $\mathbf{E}\{.\}$s relative to the $\mathbf{E}\{.\}$s of other regions of the world. It has also been documented that income is more concentrated and social protection is less effective in Latin America than in other regions of the world.[12] This, again, derives from differences in $\mathbf{E}\{.\}$, not I.H and K, or $\mathbf{T}$.

The region's $\mathbf{E}\{.\}$ results in high inequality and ineffective social protection, but they also hurt its productivity. Understanding "what is wrong" with Latin America's $\mathbf{E}\{.\}$ is critical. In the next chapter, we consider those elements associated with social protection.

3. Shortcomings of Social Protection

This chapter uses relation (5) to study the implications of social protection policies for social protection and productivity outcomes. Section one describes the formal-informal composition of the economy. Sections two to four focus on social protection outcomes: insurance and poverty and inequality, respectively. Sections five and six focus on productivity outcomes. The last two sections discuss the implications of the analysis for the rule of law and growth.

3.1. The Formal-Informal Composition of the Economy

We begin with the help of Graph 4. The first line depicts the country's human capital, I.H: a listing of all its individuals, given their schooling and abilities. Outcomes of relation (5) are shown in the next four lines. The second line describes the division of individuals between those that work and those that do not. The third divides those that work between entrepreneurs (B to C), self-employed (C to D), employed by a firm (D to E) and unemployed (E to F). Access to social insurance is shown in the fourth line, with individuals divided between informal (B to G) and formal (G to F). As illustrated by the arrows, entrepreneurs can be formal or informal, as can the self-employed or workers in firms. The last line displays individuals' income levels: profits or rents of entrepreneurs, wages of workers in firms, and earnings of the self-employed.

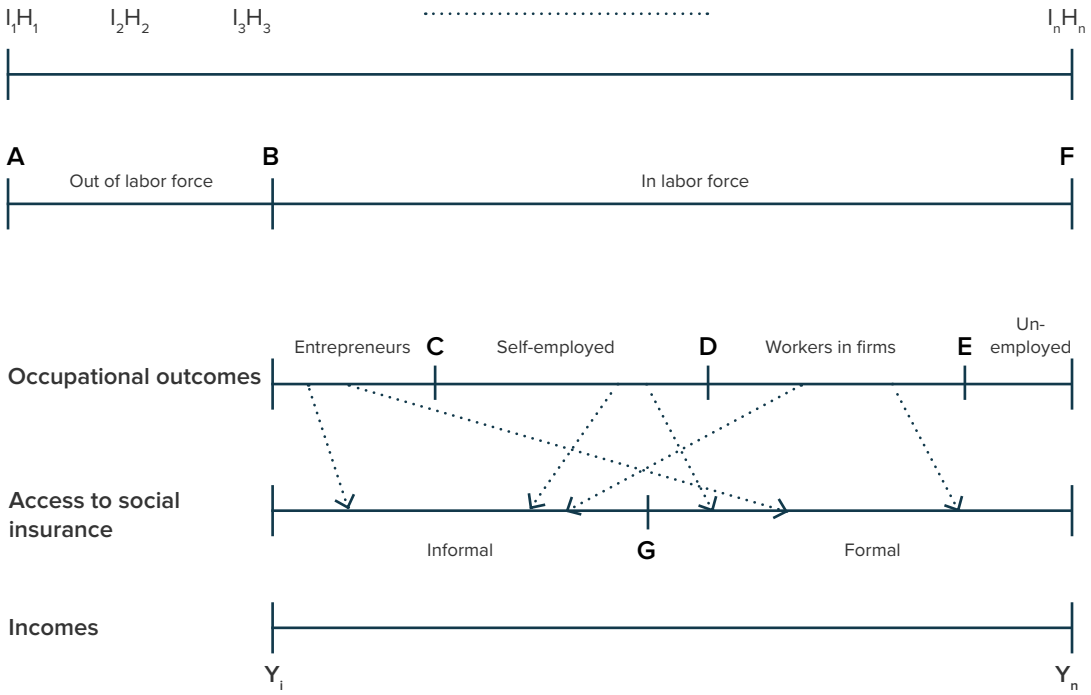
All points in the graph matter for social protection. Point B, for instance, determines the participation rate (BF/AF) and point E determines the unemployment rate (EF/BF). However, point G is particularly important because it determines the formal-informal composition of employment and thus who is insured against what risks, as formal workers are covered by contributory programs, minimum wages and protections against the loss of employment through the ($L_{CSI} + L_{JS} + L_{MW}$) regime, while informal ones receive benefits from non-contributory programs L_{NCSI}.

The location of point G varies across countries. In some, informal employment is a very large share of total employment, like in Honduras, where it is 80 percent. But even countries with

[11] See, for example, Busso et al. (2013), IDB (2010) and Fernández-Arias and Fernández-Arias (2021).
[12] Alvaredo and Gasparini (2014) show that, given its average income per capita relative to other regions, poverty in Latin America is higher. IDB (2020a) compares two measures of inequality—the Gini coefficient and the ratio of incomes of households in the 75th and 25th percentile of the distribution—between countries in Latin America, the OECD, and countries outside the region with similar development and finds that by both measures the Region is the most unequal. Lustig (2020) provides similar results.

higher per capita incomes have large shares: Colombia's share is 63 percent, Ecuador's 68 percent, Mexico's 57 percent and Peru's 77 percent. In fact, only Chile has a share below 35 percent. For the region as a whole, informal employment represents slightly over half the labor force.

Graph 4: Outcomes from Relation (5)



Source: adapted from Levy (2018).

Following the discussion in chapter two, the legal-illegal composition of informal employment varies across the region. In Argentina, all informal employment is illegal since everybody is obligated to contribute to L_{CSI} programs; the same is true in Brazil. On the other hand, in Ecuador the majority of informal employment is legal because self-employed workers are a large share of the labor force and are not required to contribute; the same occurs in the Dominican Republic. In Chile, the self-employed with Y_i under four minimum wages are also not required to contribute, so they are informal and legal; however, if their Y_i exceeds that threshold and they fail to contribute, they are informal and illegal. In Mexico, non-dependent workers in firms are explicitly excluded from ($L_{CS}I + L_{JS} + L_{MW}$) laws, as are domestic workers and the self-employed, so the majority of informal employment is legal; something similar occurs in Peru. Most countries have a mix of legal and illegal informality, with the caveat that, for the reasons explored earlier, the legal status of informal workers is sometimes ambiguous.

Tables 1 and 2 should be interpreted as outcomes of the process depicted in Graph 4. As noted, on average around 50 percent of individuals work on their own or in firms with five or fewer workers, and the size distribution of firms is strongly biased towards smallness. In parallel, most small firms are informal and have lower productivity than formal ones. The point to highlight here is that in Latin America, relation (5) results in a large share of individuals participating in economic activity as entrepreneurs, managing very small firms, including

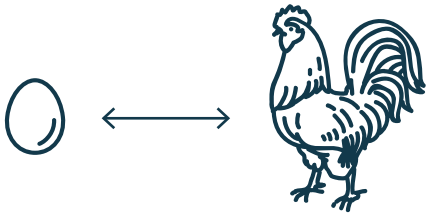
one-person firms. Clearly, something in the region's $\mathbf{E}\{.\}$s obstructs the agglomeration of individuals in larger, more productive firms and, at the same time, also obstructs the coverage of the ($L_{CSI} + L_{JS} + L_{MW}$) regime.

Graph 5 tries to shed some light on a much-debated issue in the region: the causality between informality and low productivity. The key point is that low productivity and informality are simultaneously determined by relation (5). If, for instance, some elements of T(.) or M(.) changed—the corporate tax rate or the international trade regime—some firms would attract more resources and others fewer. Depending on the TFP_is of expanding or contracting firms, **TFP** could increase or fall. Self-employment could also change. Depending on which firms offered what kinds of contracts to workers, social protection coverage would change as well. The same would happen, of course, if some elements in L(.) changed—the minimum wage or the generosity of non-contributory programs.

Graph 5: Simultaneous Determination of Productivity and Social Protection Coverage

This is the wrong question:

Is informality the result of low productivity or low productivity the result of informality?



This is the right question:

What elements in $\mathbf{E}[\ L(L_{CSI}, ..,L_{ENF}), T(T_{PIT}, .., T_{ENF}), M(M_{FIT}, ..,M_{ENF})\ ]$ result in

(i) the allocation of resources to firms with low TFP_is rather than high TFP_is, and,

(ii) incomplete coverage of the ($L_{CSI} + L_{JS} + L_{MW}$) regime?

Source: authors' elaboration.

The bottom line is that there is no sense in asserting that informality results from low productivity, or low-productivity results from informality. **TFP**—whether high or low—and informality—whether too much or too little—are both the joint result of the interactions between a country's resources I.H and K and its laws and institutions as symbolized by $\mathbf{E}\{.\}$.

From the point of view of productivity, the relevant phenomenon is that $\mathbf{E}\{.\}$ allows some firms with very low TFP_i to survive; their label is immaterial. On the other hand, from the point of view of social protection, what matters is that the insurance coverage and benefits received by workers depend on whether firms comply with the ($L_{CSI} + L_{JS} + L_{MW}$) regime; the firm's TFP_i is secondary. That said, in Latin America, surviving in the market with a low TFP_i and being outside the domain of the ($L_{CSI} + L_{JS} + L_{MW}$) regime are highly correlated; this is because, at the end of the day, the region's $\mathbf{E}\{.\}$ causes that correlation. One could conceive of situations where firms with low TFP_i are the ones that comply with the ($L_{CSI} + L_{JS} + L_{MW}$) regime; in that case, formality would be associated with low productivity. But this is not the case in Latin America.

Another important implication of relation (5) is that social protection and productivity outcomes depend on all policies in $\mathbf{E}\{.\}$, not just social protection policies, much less a single one. This makes it impossible to assert that "informal employment is caused by policy x", even less to assert that "x" is the same in all countries. Social protection policies in L(.) and T(.) are central causes of informality, but, as we discuss below, policies in M(.) also play an important role.

Further, the relevance of specific policies in L(.) or T(.) for the formal-informal composition of employment varies across countries. In Colombia, the minimum wage plays an important (though not exclusive) role, but not in Mexico. In Peru, regulations on job stability are very relevant, but less so in Chile. Informality is a complex phenomenon with many causes, and the search for "the" cause is futile. What can reasonably be done is to identify the main causes—even if this identification is not complete—and to recognize that while there are common causes across countries, their relative importance varies.

3.1.1. Formal-informal transitions

Graph 4 depicts outcomes from relation (5) at a point in time. However, firms continuously experience a mix of systemic and idiosyncratic shocks; many die and many new ones are created, only to die a few years later. Workers' preferences also change; some enter and leave the labor force, some move from one place to another, and in doing so also change from formal to informal status (or vice versa). All this implies that the same individuals can be to the right or left of point G, i.e., they can sometimes be employed formally and sometimes informally.

There are two ways to document these transitions. The first is to use employment surveys with a panel structure to follow the same individuals over a period of time. Not all countries in the region have this kind of data, but the Dominican Republic and Mexico do. In this vein, Table 3 shows one-year transitions between 2018 (rows) and 2019 (columns) across four labor statuses: formal (F), informal (I), unemployed (U) and out of the labor force (OLF).

Table 3: Transitions across Labor Status

Dominican Republic				
	F	I	U	OLF
F	78.9	9.7	3.3	8.1
I	12.5	74.0	1.9	11.6
U	17.4	30.4	29.0	23.2
OLF	5.7	14.7	1.8	77.8

Mexico				
	F	I	U	OLF
F	73.8	15.4	1.6	9.1
I	14.5	64.0	1.9	19.5
U	22.8	31.9	15.5	29.7
OLF	6.6	15.8	1.6	75.8

Note: rows add up to 100 percent.

Take the Dominican Republic. Of all formally employed workers in 2018, only 78.9 percent kept that status one year later. The rest transitioned into informal employment (9.7 percent), unemployment (3.3 percent) or left the labor force (8.1 percent). In fact, in both the Dominican Republic and in Mexico, at least 20 percent of all workers changed labor status in one year, as can be inferred by the fact that all numbers in the main diagonal (in blue) are below 80 percent. Similar numbers are observed for Brazil and Peru, which also have employment surveys with a panel structure.

Formal-informal transitions can also be inferred from the administrative registries of contributory pension programs, which record when workers are contributing, and are thus formal, and when they are not. They do so from the first time that workers enter formal employment and therefore only capture those that have been formally employed at least once. In addition, they do not record what happens when workers are not contributing, although the presumption is that some transition to informal employment, some to unemployment and some out of the labor force, as detected for the Dominican Republic and Mexico.

The administrative data are not as rich as the panel data from employment surveys, but it does shed light on workers' transitions. CAF (2020a) analyses these data in Argentina, Brazil and Ecuador, and finds that in Argentina, 3.7 percent of contributing workers fail to contribute one month later; the corresponding numbers are 2.8 percent in Brazil and 2.7 percent in Ecuador. These findings imply that over the course of one year, around 35 percent of workers in Argentina, 29 percent of workers in Brazil and 29 percent of workers in Ecuador transition from formal status to another status (informal, unemployment or out of the labor force).

The administrative data have the advantage of following workers for longer periods of time than panel data from employment surveys, which usually only do so for one year. For instance, Colombia's data allow us to follow workers between 2009 and 2019. On average during that period, 11 percent of formal workers in firms and 15 percent of formal independent workers stopped contributing each month, some because they transitioned to informal status and others because they left the labor force. In parallel, every month on average there were very similar transitions into formality. Interestingly, the data also show that, when formal, individuals also transition between self-employment and employment in firms, suggesting short stays in the same firm.

These transitions have many causes: when a firm decides to evade, its workers change from formal to informal (or when it stops evading, in which case they switch from informal to formal); when a formally employed worker decides to try their luck as self-employed (changing from formal to informal); when a micro firm goes bankrupt, and the owner-entrepreneur gets a job in a formal firm; or when workers leave the labor force. More examples could be cited.[13]

Whatever the reasons, the important point is that many workers transition between formal and informal status. In fact, one could argue that the language typically used in the region in analysis of social protection is misleading. Rather than referring to formal and informal workers—which gives the impression that that status is inherent to the person—it is more accurate to use the expressions "when workers are formally employed" and "when workers are informally employed," which reflects the fact that the adjective results from other circumstances.

[13] Agricultural workers in Mexico are an interesting example. Some are employed by firms seven months a year during the harvesting and packaging season; during the rest of the year, they are occupied by whatever occasional job comes by. These workers are formal 58 percent (=7/12) of the year, and informal the rest. In their case, transitions are driven by the rotation of the earth around the sun.

3.1.2. Formal-informal transitions by income level

Transitions are not homogeneous across income levels. In general, low-wage workers spend less time in formality than high-wage ones. The contribution density is a useful summary measure of this phenomenon, defined as the ratio of time that workers have contributed to L_{CSI} pension programs over the time that they could have contributed since they first entered formal employment. A contribution density of 100 percent implies that after a worker enters formal employment, they always keep that status. A density of 50 percent, on the other hand, signifies that the worker was formal only half of the time that they could have been, again from the first time that they acquired that status. Put differently, high densities imply that workers spend most of their time in formality, and low densities imply the opposite.

Data on contribution densities are difficult to compare across countries since they are available for different periods of time, and not always disaggregated by income level. But the cases of Chile, Colombia and Mexico are illustrative. In Chile, over the period 2009 to 2016, the average contribution density was 40 percent. But the average hides the fact that in the lowest quintile of the income distribution, 51 percent of workers had densities of at most 20 percent, and only 38 percent of these workers had densities higher than 80 percent—this compared with 9 percent and 82 percent for those in the top quintile. Results for Colombia are similar, although workers are classified by wage levels and not income quintiles. Over the period 2010 to 2020, the average contribution density was 42 percent, but was just 39 percent for workers earning up to one minimum wage, and 68 percent for those earning between 5 and 10. Data for Mexico allow workers to be followed for a longer period of time, between 1997 and 2015, but the results are the same. The contribution density increases monotonically with earnings: 13 percent for those earning one minimum wage, 39 percent for those earning two, 52 percent for those earning three and 74 percent for those earning 10 (Levy 2008; Castañon & Ferreira, 2017).

Data from CAF (2020a) for Argentina, Brazil and Ecuador show a similar picture: the contribution density increases with income. Moreover, the differences are quite significant: in Argentina, 14 percent for workers in the lowest quintile of the distribution versus 60 percent in the highest; in Brazil, 35 percent versus 71 percent; and in Ecuador, 20 percent versus 69 percent.

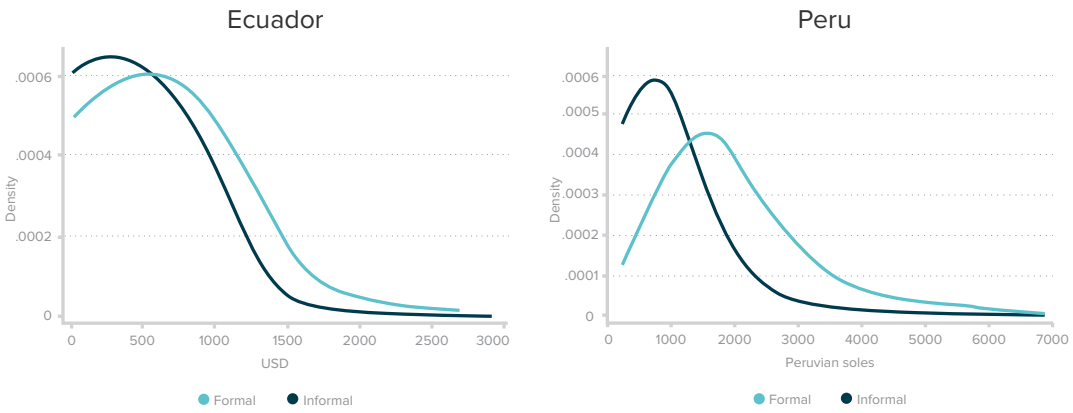
3.1.3. Overlap between the formal and informal earnings distributions

Graph 4 shows that income levels Y_i are simultaneously determined with workers' formal or informal status. Very importantly, however, the last line should not be interpreted as a ranking of workers by income level, starting with those with the lowest incomes on the extreme left and moving to those with the highest income on the extreme right. On the contrary, the evidence shows that some informal workers earn more than some formal ones.

A simple way of documenting this phenomenon is to plot the earnings distributions of formal and informal workers, as Graph 6 does for Ecuador and Peru. If all formally employed workers had higher incomes than all informally employed ones, their wage distribution (in light blue), would be completely to the right of the informal wage distribution (in dark blue), with no overlap between them. However, this is not the case, and in both countries, although the formal wage distribution is to the right of the informal one—so that the average formal wage exceeds the informal one—there is considerable overlap between them. Similar figures are

found for other countries. In all, the mean formal wage is higher, but some informal workers earn more than some formal ones. These results are consistent with the transitions between labor statuses discussed before since the same individual can be formally and informally employed. But even if there are no transitions, the data are consistent with the fact that informality is associated with many forms of labor market participation, some of which can be better remunerated than the net wages that some formal firms can pay their workers.

Graph 6: Formal and Informal Earnings Distributions



3.2. Social Protection Outcomes: Insurance

3.2.1. Limited coverage

We now discuss four implications of the formal-informal segmentation of the labor force for social insurance. The first relates to the scope of risk pooling in contributory programs. Because informal employment is widespread in most countries, this scope is substantially reduced. Since only formal workers contribute to L_{CSI} programs, the risks of illness, death or disability are pooled only between them. This is relevant because the costs of insurance are lower when the pool of participants is more numerous and diversified. Smaller pools mean higher costs.

The second relates to protection against the loss of employment. As mentioned, in many countries, job stability regulations substitute for unemployment insurance. But they are a poor substitute. These regulations try to prevent unemployment by making it very difficult for firms to dismiss workers, rather than helping workers when they are unemployed. Firms react by limiting the number of workers hired formally. Those who get a formal job with a permanent contract are indeed well protected against negative shocks, but the rest are not protected at all. Rather than being pooled, risks from negative employment shocks are transferred to informal workers.[14]

[14] That said, even risk pooling among formal workers may be reduced, as in Chile, where contributions to unemployment insurance are channeled to workers' individual accounts. When they lose their job, benefits are first drawn from these accounts and only when they are exhausted, from a common fund, so that in large measure unemployment risks are covered through self-insurance.

The third implication relates to savings for retirement pensions. Regardless of whether modalities are pay-as-you-go (PAYG) or individual accounts, workers are required to save only when they are formally employed. Because hardly any informally employed workers voluntarily save through the contributory pension regime, aggregate pension savings are lower. This is very relevant because in most countries these savings are the largest source of long-term savings in the national currency. Reduced coverage of L_{CSI} limits the resources available for long-term investment projects.

Finally, the fourth is that the formal-informal segmentation makes it difficult to respond effectively to large and unexpected systemic shocks. In general, informal workers voluntarily access whatever non-contributory programs are available (although there may be capacity constraints), but without any obligation to register with social security institutes (and, in some countries, tax authorities). This makes it difficult to identify and reach them when shocks occur that require a rapid response (an issue that Covid-19 brought to the fore).

With varying degrees of coverage across countries, when workers are informally employed, they get some insurance through non-contributory programs. But, with a few exceptions, these programs insure only against illness and longevity risks (through non-contributory health and retirement pension programs). The vast majority of informal workers in the region is not insured against disability, death or accidents at work—risks that are always covered in L_{CSI}. Further, no country includes protections against the loss of employment in L_{NCSI}. Undoubtedly, workers are much better protected against risks when formally employed.

Thus, as opposed to the OECD, in Latin America social insurance is only partly social, as *de facto* many households do not participate systematically in the mechanisms for risk pooling and savings for retirement that are associated with such insurance, nor are they permanently protected from very relevant contingencies. In Latin America, households bear many more risks on their own, or through their network of family and friends, than in OECD countries, where these risks are more broadly shared. One could say that the formal-informal segmentation of the labor force constrains solidarity across members of society, narrowing the space for nourishing social capital. These are all grave shortcomings. Unfortunately, there is more.

3.2.2. Erratic coverage

Formal-informal transitions significantly reduce the efficacy of insurance. Consider first health. In the region, the quality of services of contributory programs is generally better than that of non-contributory ones; more complex and costly interventions are covered, wait times are shorter and there is better availability of medicines. As a result, as workers transition across statuses, they (and their families) receive different quality of care. This difference matters for out-of-pocket expenditures for medicines when there are minor complications like influenza or stomach infections. But it matters more if a serious problem like cancer or kidney failure occurs. Since these ailments are rarely covered by non-contributory programs, when workers are informally employed, they have to pay for treatments with their own resources, and these potentially catastrophic expenditures can leave them in financial ruin. Further, contributory services are usually provided by social security institutions while non-contributory services are provided by health ministries, each with their own facilities, protocols and records (although in some countries like Argentina, Chile and Colombia provision can be subcontracted to

private institutions). In this context, formal-informal transitions result in interrupted treatments and follow-ups for chronic diseases like diabetes or high blood pressure.[15]

Consider next retirement pensions. Many countries require workers to contribute for a minimum number of years to be entitled to one, regardless of whether they are defined contribution or PAYG. In Brazil, Ecuador and Honduras, they must contribute 780 weeks; in Peru, 1,000; in Mexico, 1,250; in Colombia, 1,300; and in the Dominican Republic, 1,500. In Argentina, workers are also required to contribute 1,500 weeks, although this requirement is lowered through a host of special regimes.

Transitions make it impossible for many workers to reach that threshold. These workers, sometimes formal and sometimes informal, contribute to a pension during part of their working life, but are denied one at retirement. This is a serious issue, as exemplified by these statistics:

- in Colombia, between 65 and 75 percent of workers that contribute to a pension will not get one, depending on whether they are in the PAYG or individual accounts system;
- in Argentina, Ecuador and Honduras, less than 50 percent of workers participating in the PAYG system will get a pension;
- In Mexico, two out of every three workers contributing for a pension through the defined contribution modality will not qualify for one; and,
- in Peru, 60 percent of workers contributing to its PAYG system will not get a pension.[16]

But even if we ignore thresholds, transitions imply lower pensions since accumulation occurs only during formal status. Put differently, contribution densities are low. In Peru, for instance, the average contribution density is 36 percent; in Mexico, 44 percent, in Colombia, 45 percent and in the Dominican Republic, 47 percent. This compares with an OECD average of 59 percent.

The combination of many workers being left without a contributory pension even if they save for one, together with low replacement rates for the majority of those who do get one, leads to an unambiguous conclusion: in Latin America, these pensions are deeply flawed.

Countries in Latin America have a mix of retirement pension systems: defined benefit PAYG and defined contribution through individual accounts. Brazil and Ecuador have the former and Chile and Mexico, the latter. Some countries, like Colombia and Peru, have both and allow workers to choose between them; other countries, like Honduras, obligate workers to contribute to both at the same time if their earnings exceed one minimum wage. Argentina adopted a defined contribution individual accounts system in the mid-1990s, which co-existed alongside a publicly managed PAYG system, and then nationalized the defined contribution system in 2008.

[15] Doubova et al. (2018) find that in Mexico, over three years, 32 percent of patients with Type II diabetes lose their right to be treated in the social security institute, leading to a 43 percent drop in the quality of treatments and a 19 percent decline in clinical outcomes.

[16] In Peru, even workers that qualify for a pension may not get one because they can dispose of their savings when they reach retirement age. Since 95 percent of those retiring do so, de facto Peru's defined contribution system does not generate pensions, only savings for the day of retirement, fully passing on the risks of longevity to workers.

There has been a long-standing debate about relative merits of each system.[17] But formal-informal transitions and the ensuing low contribution densities imply that this debate, while important, is secondary. No system of retirement pensions is going to work well when workers only save for their pension for part of their working lives. As things stand today in the region, formal-informal transitions severely limit the number of workers who will obtain a pension and reduce its amount for those who will get one—never mind if the system is PAYG or defined contribution.

Transitions also reduce the efficacy of unemployment insurance for the few countries that have it. Regulations require workers to be formally, continuously employed for two or three years to qualify. Since many do not meet this threshold because of transitions, when dismissed they may not be protected from income loss even if they were holding a formal job.

More generally, under formal-informal transitions, insurance against risks is erratic. If a worker dies while formal, their family gets a survivorship pension, but not if he dies while informal; if a worker delivers a baby while formal, she gets maternity pay and leave, but not if that occurs when she is informal, and so on. It is as if one's house were insured against fires or earthquakes in the spring and fall, but not in the summer and winter.

Workers in all countries transition from job to job throughout their lifetime. Their preferences change: at times they want to work for themselves and at times for a firm. In parallel, firms close, start, grow and contract. Everywhere there is large firm and worker churn. But in most OECD countries, job transitions have no impact on workers' access to social insurance because the system is not segmented. However, in Latin America, job transitions are often associated with changes in labor status, with major implications for workers' protections against risks.

3.3 Social Protection Outcomes: Poverty

This section focuses on the implications of relation (5) for poverty. The departure point is the last line in Graph 4 listing individuals' incomes $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$. As noted, the overlap between the formal and informal earnings distributions is substantial, as shown for Ecuador and Peru (and for Mexico in Graph 8). It is true that on average informal incomes are lower than formal ones, but this does not imply that all informally employed individuals earn less than formally employed ones. It implies even less that informality should be equated with poverty.

To discuss poverty, we need a cut-off level of income, $\mathbf{Y}^{cutoff}$, separating the non-poor, ($Y_i > \mathbf{Y}^{cutoff}$), from the poor, ($Y_i < \mathbf{Y}^{cutoff}$). We can then calculate a poverty index, labelled here $P(\mathbf{Y}, \mathbf{Y}^{cutoff})$, which can simply be the share of poor individuals, or also incorporate information about the gap between their income Y_i and $\mathbf{Y}^{cutoff}$ (like a Foster-Greer-Thorbecke index).

[17] There are many differences between them, but a key one is the management of longevity risk. In PAYG systems, it is pooled among all workers by the social security institute. In defined contribution ones, it is transferred to private insurance companies through the purchase of annuities. That said, the region's experience with the second system has not been a happy one because administration fees are high, reducing the size of the pension, and because annuity markets are very thin. On the other hand, PAYG systems have been characterized by actuarial imbalances between contributions and benefits, resulting in large subsidies that heavily burden the fiscal accounts.

Multidimensional indices that also consider access to health and social insurance can also be calculated from Graph 4. But for our purposes, a simple head-count measure suffices.

There is a large literature on how to measure $\mathbf{Y}^{cutoff}$, which we do not discuss here. But whatever the measurement method, its value matters substantially. It determines the location of the horizontal dotted line in Graph 2, and thus the number of workers that fall in quadrants III and IV in principle entitled to targeted poverty programs.

Table 4 shows poverty rates for two values of $\mathbf{Y}^{cutoff}$, US$1.90 and US$5.50 dollars a day, as well as the labor informality rate for various countries.

Table 4: Poverty and Labor Informality Rates

	Y^{cutoff} = US$1.90	Y^{cutoff} = US$5.50	Informality (%)
Argentina	0.7	7.8	36
Brazil	4.1	20.7	37
Chile	0.8	7.9	35
Colombia	4.5	28.3	63
Dominican Republic	1.6	21	56
Ecuador	3.6	24	68
Honduras	16.0	50.5	82
Mexico	2.2	25.7	57
Peru	3.5	24.4	77

Source: Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank). Poverty lines based on purchasing power exchange rates for 2011. Poverty rates for 206-17, except Chile 2015.

Even if the higher cut-off is used, it is clear that poverty and informality are two distinct phenomena. In fact, with the exception of Honduras, in all countries there are more non-poor than poor informal workers (or, in terms of Graph 2, more workers in quadrant II than IV). What needs to be highlighted is that in Latin America there are many workers whose incomes exceed $\mathbf{Y}^{cutoff}$, but who lack the protections derived from the (L_{CSI} + L_{JS} + L_{MW}) regime. That said, most poor workers are informal, so they are only insured against whatever risks are covered by L_{NCSI}.

Graph 7 provides further information for the Dominican Republic and Honduras, classifying the labor force (excluding the unemployed) by labor status and income levels, following Graph 2. In this case, however, poverty is measured with the criteria applied by each country, not the values of $\mathbf{Y}^{cutoff}$ used in Table 4. Even if a country's poverty measurements are used, the same result holds: poverty and informality are not the same. In the Dominican Republic, there are more non-poor than poor informal workers; in Honduras the opposite holds, but even there, 30 percent of workers are informal but not poor.

Graph 7: Poverty and Informality are Different

Dominican Republic

	Formal	Informal
Non-poor	36%	44%
Poor	6%	14%

Honduras

	Formal	Informal
Non-poor	13%	30%
Poor	4%	52%

3.3.1 Income transfers

Policies to help the poor in Latin America can be grouped in two sets. The first are income transfers through various means: exemptions to consumption taxes (like lower VAT rates for food) or subsidies to transportation, gas or electricity. These transfers can be generalized or targeted but are usually disassociated from any behavior by households (other than consuming whatever is subsidized). When they are generalized, like sometimes occurs with gasoline and other energy products, they benefit all households. This merits attention because the size of the transfer received by households is proportional to their income, meaning that in absolute terms non-poor households receive larger transfers than poor ones (IDB, 2019). Because the poor consume a smaller share of subsidized goods than the rich, generalized subsidies are inefficient mechanisms to transfer income to them.

The second set of policies are targeted transfers, which can be in-kind or in cash. The first includes school breakfasts, distribution of milk or other staples, and the like; the second includes monetary transfers, which are often conditioned on investments by households in their human capital. These programs are better known as CCTs, a Latin American innovation started in the mid-1990s in Mexico, then in Brazil and later adopted in Colombia, Ecuador, Honduras and many other countries (including outside the region).[18]

Importantly, resources for CCTs or other targeted transfers are often less than those channeled to exemptions to consumption taxes or generalized subsidies, sometimes by a significant margin. Table 5 contrasts government spending on generalized subsidies for energy, as calculated by the IDB (2017b), with cash transfer programs targeted to the poor (CCTs or non-contributory pensions) for a sample of countries, in both cases expressed as a share of GDP. Because energy subsidies fluctuate depending on the global oil price, the figures are the average for each country between 2008 and 2014. Clearly, there is a large imbalance between generalized and targeted transfers, with the former being several times greater than the latter (a situation that would be aggravated if tax exemptions were considered).

Table 5: Energy Subsidies and Targeted Transfers for the Poor

	Argentina	Colombia	Ecuador	Honduras	Mexico	Peru
Energy	3.0	1.4	6.4	1.9	1.9	1.0
Transfers	0.6*	0.31	0.8	0.2	0.4*	0.22

* Excludes non-contributory pensions since they are not targeted on the poor.

This unbalance, combined with the fact that targeted programs sometimes have significant exclusion errors, result in many poor households receiving little income support.[19] On one hand, their purchasing power is low and therefore they benefit little from generalized subsidies; on the other, at times they are excluded from CCTs or other targeted programs. This partly explains why poverty rates in the region are high relative to its income level.

[18] There is by now a large literature on many aspects of these programs; a few examples are Alzua et al. (2012) De Brauw (2015b), Fizbein and Schady (2009), Garganta et al. (2017), Levy (2006), Saavedra and Caballero (2018), and Skoufias (2008).

[19] For instance, in the Dominican Republic, only four out of ten households entitled to their CCT (Progresando con Solidaridad) actually receive it.

Targeted transfers have an inevitable trade-off. From a fiscal point of view, they are substantially less costly than generalized transfers. If only 30 percent of the population is poor, there seems to be little point in giving subsidies to the other 70 percent who do not need them. However, any targeting scheme involves an implicit tax since if poor households earn additional income that places them above $\mathbf{Y}^{cutoff}$, they lose the transfer. The net income from their efforts is then the additional income earned less the value of the foregone transfer. If the transfer is high, the implicit tax is high, and poor households may be caught in a poverty trap; see Besley and Kanbur (1990). This is an important issue, which speaks to the potential pitfalls of overreliance on targeted transfers and argues for a better balance between these transfers and other policies to help the poor. We return to this discussion in chapter four.

3.3.2. Income transfers versus insurance

Income transfers do not substitute for insurance. Like everyone else, poor workers experience idiosyncratic shocks: they can get seriously ill, lose their job, die or suffer a permanent disability. The income transfers that they can obtain through targeted programs, when they get them, are not effective to protect them against these shocks.

We make two points. First, poor households need both income transfers and insurance, not one or the other, and second, poverty measures like $P(\mathbf{Y},\mathbf{Y}^{cutoff})$ fail to capture the impact of the increased security obtained from being covered by L_{CSI} programs on the welfare of poor households, even if this coverage is not reflected in their monetary income. Put differently, a household with Y_i slightly below $\mathbf{Y}^{cutoff}$ but covered by L_{CSI} programs (in quadrant III of Graph 2) would in all likelihood be better off than one whose Y_i is slightly above $\mathbf{Y}^{cutoff}$ but is excluded from L_{CSI} programs (in quadrant II).[20]

The region's poverty rate fell from 42 percent in 2002 to 23 percent in 2018, as calculated by the IDB (2020a) using US$5.50 a day as $\mathbf{Y}^{cutoff}$. This has increased the size of the middle class. However, an important share of this middle class is in a vulnerable position. IDB (2020a) estimates that about 37 percent of all households in the region have incomes between US$5.50 and US$13 a day. Some of them enter and exit poverty regularly (Vakis et al., 2016). Living at the edge of poverty, a positive idiosyncratic shock can place them above that edge, but a negative one can push them in the opposite direction. Cruces et al. (2011) find that, in a fifteen-year period, 10 percent of all people in Latin America fall into poverty every year period (See also López-Calva et al., 2014). Lack of insurance is one of the reasons.

This issue is particularly relevant in countries, like Colombia, that restrict their non-contributory health programs only to poor workers. When a serious illness occurs, non-poor households can easily fall into poverty. Alternatively, when a systemic shock occurs, as with Covid-19, the insurance provided by the family network is rapidly exhausted since all are facing the same hardships at the same time; thus, they all fall into poverty. For these households, insurance is key. Being left out of the protections provided by the $(L_{CSI} + L_{JS} + L_{MW})$ regime is a major problem, and targeted transfers are a poor solution. Rather than acting *ex ante* to prevent

[20] Molina Millán et al. (2019) show that the reduction in uncertainty about future income associated with CCT programs increases households' welfare by reducing stress. One can expect the same type of effects, perhaps stronger, if households are given certainty about the coverage of other risks.

poverty, policies react *ex post* to ameliorate it once it is here. Poor households are helped to endure poverty, but not to prevent it; in other words, there is some support for those at the bottom, but no insurance to avoid the fall.

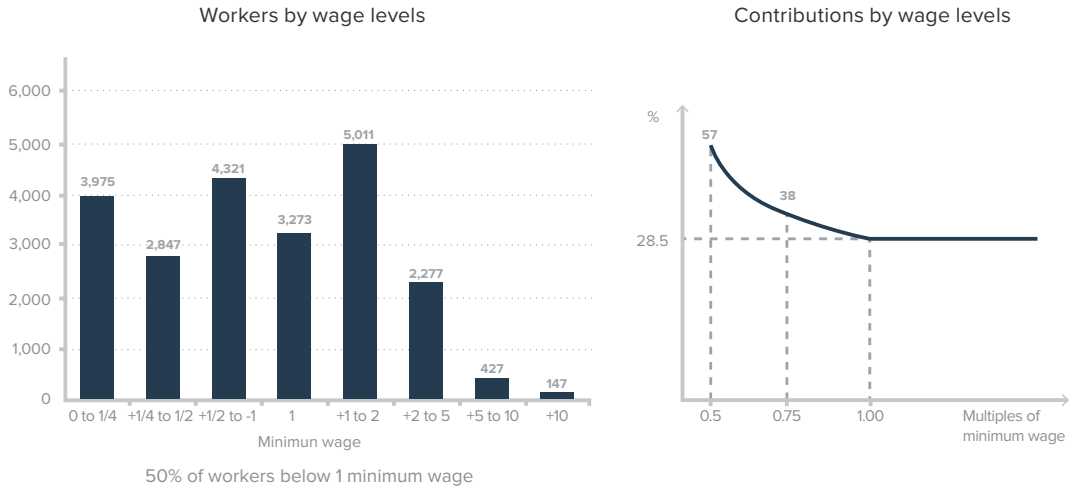
3.3.3. Interphase between insurance, minimum wages and poverty programs

Income transfers for the poor should not be conditioned on their labor status. Unfortunately, in some countries they are. In Ecuador, households covered by its CCT (*Bono de Desarrollo Humano*) lose their benefits if they get a formal job, as if being insured by L_{CSI} programs eliminates their need for income support; the same occurs in Peru (*Juntos*). Unsurprisingly, this increases informality rates for poor workers, as found for women in Ecuador by Araujo et al. (2017) and Bosch and Schady (2019). Argentina's CCT (*Asignación Universal por Hijo*) was initially only compatible with informal work, unemployment (though not from a formal job) or non-participation in formal employment, reinforcing incentives to be informal (Garganta & Gasparini, 2015).

The interphase between L_{CSI} regulations and the minimum wage can also deter poor workers' entry into formality, even if CCTs are in principle neutral to their labor status. In some countries like Ecuador, Honduras, Mexico and Peru, the self-employed are not required to contribute to L_{CSI} programs but can do so voluntarily. However, contributions are based on at least the minimum wage, a relevant consideration since many self-employed workers in these countries earn less than that. This is partly because the minimum wage has increased faster than their earnings, like in Ecuador, where it tripled in real terms between 2000 and 2019.

Graph 8 illustrates this issue in Colombia, where contributions for L_{CSI} programs for the self-employed are 28.5 percent of wages: 12.5 percent for health and 16 percent for pensions. The left panel shows the number of workers in each wage range and indicates that half of the labor force earns less than one minimum wage. The right panel shows the implied contribution rates as a function of wages, considering that they are based on at least the minimum wage; for those earning one-half of that, the contribution rate is 57 percent, and for those earning three-fourths, 38 percent. In other words, if these workers wanted to be formal, they would need to devote at least a third of their earnings for health and pensions. On its own, this would seem a sufficient reason for them not to contribute. But two factors strengthen that decision: one, if they do not contribute, they have free access to health benefits of the same quality. Two, if they do, they need to do so for 22 years to qualify for a pension under the defined contribution modality or 25 years under the PAYG modality, a very unlikely event.

Graph 8: Contributory Insurance, Minimum Wages and Self-employment in Colombia



The alternative route that poor self-employed workers in Colombia have to be formal is to be hired by a firm as dependent workers. But in that case, the firm has to pay the minimum wage, contributions to L_{CSI} programs—which cannot be shifted back to workers in the form of lower wages because of the minimum wage—and bear the contingent costs of job stability regulations. Because this is not profitable for most firms, the result is that most poor workers are self-employed, or work in micro firms that can either elude the obligation to contribute because the relation of dependency is ambiguous or evade that obligation by taking advantage of weak enforcement. Whatever the case, most are informal.

In Brazil, poor workers do not automatically lose their CCT (*Bolsa Familia*) if they get a formal job, but the program's rules imply that they probably would. A formal job must pay at least the minimum wage. That wage may place the household's per capita income above the threshold used to select households into *Bolsa Familia* (about one-fourth of a minimum wage), particularly if it has few members. In addition, poor workers may also be reluctant to get a formal job if they consider it to be short-lived since getting back into *Bolsa Familia* is difficult. These factors, in the context of important increases in the real minimum wage, help to explain why formality rates among beneficiaries fell from 10 percent in 2004, the year after the program started, to 6.4 percent in 2019.[21]

More generally, conditioning access to poverty programs on informal status, conditioning access to non-contributory programs on being poor, or using the minimum wage as a basis for contributory programs for the self-employed, are all bad ideas; *de facto*, large implicit taxes are imposed on improving one's status, as Bérgolo and Cruces (2021) find in Uruguay. This is an important issue to which we return below, but we note that, sadly, these ideas permeate regulations on targeted poverty programs and non-contributory programs in many countries.

[21] De Brauw et al. (2015a) show that Bolsa Familia does not influence overall workforce hours of work. However, they do find a shift of nearly eight hours of work per week per household member from the formal to the informal sector.

3.3.4. Permanent versus transitory transfers

Income transfers through targeted poverty programs should be transitory, supporting households while they work their way out of poverty. Ideally, poverty should end. Insurance, on the other hand, needs to be permanent: households can suffer negative shocks any time and regardless of income levels. But in the absence of insurance, targeted income transfers become a permanent necessity. The conflation between income transfers and insurance, and the poor quality of insurance provided by non-contributory programs, is another reason why poverty persists in Latin America, in addition to the large imbalance in the allocation of budgetary resources between targeted and untargeted transfers, and the imperfect coverage of targeted programs.

There is a third reason. CCTs have a dual purpose: to raise poor households' current consumption while investing in their human capital and to break the intergenerational transmission of poverty. The evidence indicates that by and large they have been successful in the first objective: consumption increased, undernutrition fell, morbidity decreased and schooling indicators like attendance and progression rates improved.[22] But the same cannot be said of the second. Improvements in the human capital of children and youth have not translated fully into higher-earning jobs after they enter the labor market. Some earn higher incomes, but many are occupied in the same low-productivity, low-income informal activities as their parents; their increased consumption results from the transfer, not from higher earnings associated with their improved human capital.

Section 3.5 elaborates on the reasons behind this outcome, but the general point can be made here. The incomes of the poor are not solely determined by an individual poverty program like a CCT, effective as this program may be. They are determined by all policies in $\mathbf{E}\{.\}$ and, prominently, other social protection policies in L(.) and T(.) aside from CCTs. And these policies have *de facto* made it very difficult for poor workers to obtain higher-productivity formal jobs.

With the benefit of hindsight, it is clear that since the early 1990s, while much of the region's attention was centered on improving the human capital of poor children and youth in the hope that they would get better jobs when they entered the labor force, insufficient attention was focused on which firms were going to employ them. Would they be firms that required their enhanced human capital, offered them long-lived contracts and invested in their training? Or would they be the same small, low-productivity informal firms that employed their parents? The numbers in Table 1 are relevant to this discussion. The majority of the poor are self-employed or work in very small and fragile firms. They are there because of their countries' $\mathbf{E}\{.\}$s, not because they lack human capital.

Permanently subsidizing gas and electricity, and exempting some consumption goods from VAT, helps the non-poor more than the poor. The root problem—low earnings due to a lack of productive jobs—is not solved by increasing generalized consumption subsidies. Increasing targeted transfers may, for the reasons discussed in section 3.5, worsen this problem.

It is useful to return to Graph 2. Poverty policy at times takes the target population of its programs, quadrants III and IV, as exogenously given. But the number of workers in these quadrants is not exogenous; it depends on social protection policies. As discussed below,

[22] For a region-wide review of the evidence, see Fiszbein and Schady (2009) and Molina Millán et al. (2019); for Argentina, Garganta et al. (2017); for Brazil, De Brauw et al. (2015b); for Colombia, Baez and Camacho (2011); for Honduras, Molina Millán et al. (2020); and for Mexico, Parker and Todd (2017).

if the (L_{CSI} + L_{JS} + L_{MW}) regime worked better, formality and incomes would increase and the need for targeted poverty programs would diminish. Put differently, at some point, improving (L_{CSI} + L_{JS} + L_{MW}) programs is a more effective policy to reduce poverty than increasing the generosity of targeted transfers.

We should no longer expect targeted income transfers for the poor to eliminate poverty on their own. This is not to say that they should end; they should not. But it is to say that, given the targeted programs that are already in place in most countries, what is most needed to help the poor are higher-productivity jobs with access to proper insurance. This objective will only be achieved if poverty policy expands its perspective beyond individual targeted programs and considers the whole range of social protection policies, particularly those that affect the performance of the labor market. If this is not so, and the belief persists that poverty can be eliminated through self-contained targeted programs, Latin America will continue to trap a significant share of its population in poverty, investments in their human capital notwithstanding.

3.4 Social Protection Outcomes: Inequality

We turn to the main implications of relation (5) for income inequality. The starting point is again the list of incomes **Y** = (Y_1, Y_2, ..., Y_n) in the last line of Graph 4. The most common measure of inequality over those incomes is the Gini coefficient, denoted G(**Y**), and it is the one that we use here. In 2017, its value was 0.41 for Argentina, 0.53 for Brazil, 0.46 for Chile, 0.50 for Colombia, 0.46 for the Dominican Republic, 0.45 for Ecuador, 0.50 for Honduras, 0.48 for Mexico and 0.43 for Peru.[23] These numbers are high compared to countries in the OECD, indicating higher inequality in the region; see also IDB (2020a) and Lustig (2020).

Relation (5) highlights that incomes Y_1, Y_2, ..., Y_n and therefore G(**Y**), result from all policies in $\mathbf{E}\{.\}$, not only social protection ones. This observation matters because some manifestations of inequality, particularly extreme concentration at the high end of the distribution (the "top 1 percent"), may be associated with failures in the laws and institutions in the "realm of market conditions" M(.), particularly those designed to combat monopolies and abuses of market power. This results in large monopoly rents or failures in policies to combat corruption. These failures redistribute from all households to those at the top, the opposite of what is desired.

Social protection policies cannot correct these failures. If a rich monopolist extracts 100 pesos of rents from all other households, even a high marginal rate of the personal income tax, say 50 percent, only corrects half of the problem. On the other hand, if competition policies are effective, there are no monopoly rents. In that case, income inequality is less even though the ex-monopolist pays less in taxes. Put differently, effective policies in M(.), are indispensable elements of the region's redistributive toolkit.

---

[23] These estimates are obtained from household surveys. However, many researchers have pointed out that these surveys fail to properly capture the incomes of households at the very top of the distribution (Alvaredo et al., 2018). In some cases, it is possible to complement them with data from tax and other administrative registries, and this can make a substantive difference. For instance, Flores et al. (2019) find that in Chile using household surveys only, the share of income captured by the top 1 percent fell from 18 to 13 percent between 1990 and 2015. However, when these surveys are complemented with administrative data, they find that that share was initially higher, 22 percent, and only fell to 21 percent. We do not pursue this issue further as our purposes here are different, but it is important to keep it in mind when interpreting data on levels and trends on inequality derived only from household surveys.

This issue merits attention because it transcends its impact on G(**Y**). Extreme concentrations at the top give great power to a few individuals, who use it to influence the political process to change laws and regulations in $\mathbf{E}\{.\}$ in their favor or, at least, to change how these laws are applied to them—a situation that can be described as one where all are equal under the law, but some are more equal than others.[24] This situation results in perceptions of privilege and unfairness of the system. Undue exercise of influence and special privileges cannot be remedied by social protection policies. These vices result from other flawed policies in $\mathbf{E}\{.\}$, and they must be corrected not only to lower G(**Y**), but, more importantly, to improve the functioning of the region's democratic processes.

3.4.1. Contributory social insurance and inequality: incidence of contributions

Social protection policies also impact individuals' incomes and thus G(**Y**). We discuss first those that provide insurance through contributory programs. The expectation is that they reduce G(**Y**) through two complementary channels:

- firms pay a share of the contributions; and,
- some programs involve cross-subsidies from high- to low-income formal workers.

Begin with the first channel. Laws on social insurance state that firms and workers contribute to L_{CSI}, with variations in the share paid by each across health, pension and other programs; see Alaimo et al. (2017). The contributions paid by firms reduce G(**Y**) under two assumptions:

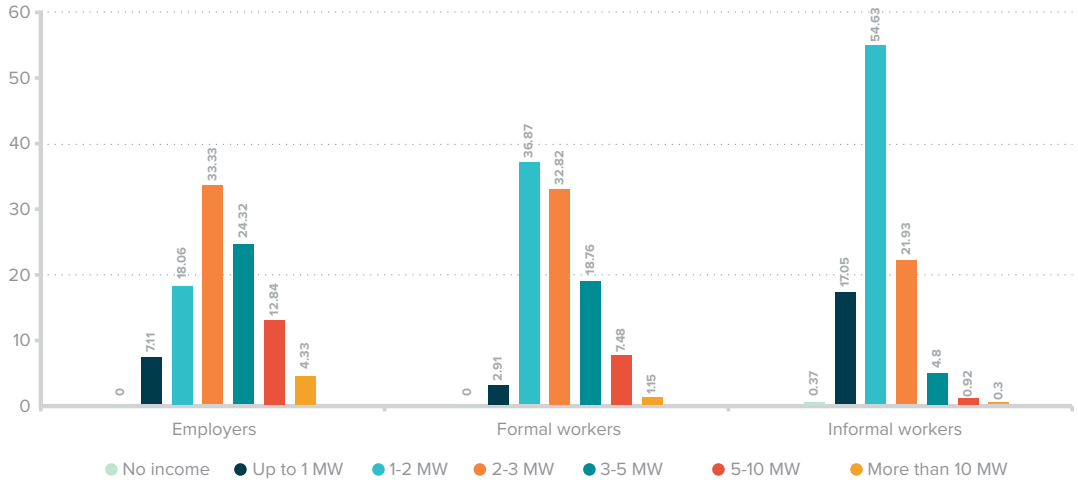
- firm owners have higher incomes than workers; and,
- firms absorb their contributions in the form of lower profits.

Although the first assumption is rarely questioned, it is useful to recall three facts about firms in the Latin America: the border between a self-employed worker and a micro firm is very fluid; the vast majority of firms are informal and small; and, in many of these firms, owners and workers are often relatives of each other. These facts imply that the border between workers' remunerations and firms' profits is often fuzzy. The surplus of the firm is distributed among participants through rent-sharing agreements or cultural norms. It is hard in these contexts to establish that all firm owners have higher incomes than workers.

Graph 9 sheds some light on this issue in Mexico, where the employment survey separately identifies the incomes of employers and workers and records the size of employers' firms. Incomes are measured in multiples of the minimum wage. As expected, the average for employers is the highest, 3.8, versus 2.8 for formal workers and 1.7 for informal ones. The more important observation, however, is that the distributions overlap. As it happens, 27 percent of formal workers have an income of three or more minimum wages, while 58 percent of employers have incomes below that threshold; put differently, some formal workers earn more than some employers. (Note also the overlap between the formal and informal wage distributions, similar to the results for Ecuador and Peru shown in Graph 6.)

[24] See the discussion in Chapter 3 of the 2021 Latin American and Caribbean UNDP report.

Graph 9: Incomes of Employers and Workers in Mexico



These results are not surprising in light of the data on firms presented in Table 2. Most employers manage very small firms. In the Mexican data, those earning up to three minimum wages manage firms with 1.9 workers on average; those earning more than 10 minimum wages manage firms with 12.1 workers on average, but they are a small minority.

Unfortunately, the Mexican data are insufficient to establish whether employers own firms. For small firms one could suppose that generally this is so since they are mostly informal and some are not even incorporated as a separate legal entity.[25] It is difficult to think that the owner of a two-worker firm hires (and pays) a third person to manage it. If employers and owners in these firms are the same, the result is that many owners have income levels that are not too dissimilar from those of their workers. In these cases, the assumption that firm contributions to L_{CSI} programs redistribute income from high-income firm owners to low-income workers does not apply. In fact, following Table 2, the assumption probably does not apply to the majority of firms in the Latin America.

Of course, not all firms are small and informal or have a contractual structure where owners and workers are relatives, or where employers are owners. There are firms where there is a clear distinction between the firm owners and their workers, and where in all likelihood the income of the former exceeds, often by large orders of magnitude, that of the latter. These firms are mostly formal and, in relative terms, substantially larger than informal ones. And although they are a minority, they account for an important share of employment in the region, between 20 and 30 percent depending on the country and threshold chosen. For these firms, the assumption that owners have higher income than workers surely holds.

The point of this discussion is that—contrary to what is often implicitly assumed—the assertion that all firm owners in Latin America have higher incomes than workers is unwarranted; the situation is more nuanced. Of course, if only formal firms are considered, the assertion is more likely to be true. In these cases, redistribution from firms to workers through contributions

[25] Over 95 percent of informal firms with five or fewer workers, and more than half with 6 to 10, are not incorporated (Levy, 2018).

to L_{CSI} is equivalent to redistribution from high- to low-income households, which is what is required to lower G(**Y**).

That said, it is of the essence to keep in mind that the vast majority of firms in the region are very small. Consider again Ecuador in Table 1. There are nearly three million firms, almost all of which have between 2 and 10 workers. These firms employ 5.7 million workers, for an average firm size of 1.9. Only about 11,000 firms have 11 or more workers, employing 1.5 million workers, for an average firm size of 135. Unfortunately, there is no data to compare the income of firm owners versus workers, but it stands to reason that in the second set of firms, these differences are substantially larger than in the first. In this context, the impact of L_{CSI} contributions on G(**Y**) in Ecuador is ambiguous.

What about the second assumption? The issue here is that while the letter of the law might state that a share of L_{CSI} contributions is paid by firms, real incidence depends on the functioning of the labor market.

To see this, assume that $L_{CSI} = 0$. In this case, what formal firms pay for labor is exactly what workers receive. Let this be the wage w*. Assume now that $L_{CSI} > 0$, and that the law states that contributions are paid by firms. For this to be so, workers' wages have to remain at w*, implying that firms' labor costs are now $(w^* + L_{CSI})$. However, firms will react to higher labor costs through various strategies. Some will simply reduce formal employment. Others may decide to cheat and hire some workers informally, to elude and give some workers a contract masking the relation of dependency and subordination, or to subcontract some tasks to other firms. The mix of adjustment strategies depends on the shape of firms' demand functions for labor, on the creativity of its lawyers, and on enforcement. But what will most likely not happen is that all firms hire the same number of formal workers when $L_{CSI} > 0$ as when $L_{CSI} = 0$ at the wage w*.

As formal employment falls, the same happens to formal wages, which are now $w^{**} < w^*$. In the end, despite the letter of the law, firms pay $(w^{**} + L_{CSI})$, not $(w^* + L_{CSI})$; the difference is paid by workers in the form of lower wages, as a share of L_{CSI} contributions is shifted back to them.

Many factors determine that share. An important one, discussed further in section 3.5, relates to whether workers consider that one peso of contributions is worth one peso of benefits to them. If that is so, and if there are no impediments to changes in wages like minimum wages, then that share equals one. In other words, there is "full shifting" so that $w^{**} = (w^* - L_{CSI})$ and there is no change in employment because firms' labor costs are the same. But these two conditions are rarely met, and full shifting is unlikely; in general, the adjustment to L_{CSI} occurs partly through lower wages, partly through lower firm profits, and partly through lower employment.[26]

Backward shifting dilutes the redistributive impact of L_{CSI} contributions, even assuming that firm owners have higher incomes than their formally employed workers. The magnitude of the shift depends on workers' valuation of L_{CSI} programs and characteristics of countries' labor markets, and therefore varies across the region. For instance, the minimum wage may

[26] See Summers (1989) for the classic analysis and Levy (2008) for an extension to contexts with informal employment. Note that if full shifting occurs, there is no change in employment, but also no redistribution. Bozio et al. (2020) review various studies on incidence showing that shifting is higher when there is a stronger relationship between contributions and benefits.

set a floor on the fall in wages, because w** must be at least equal to it. In countries where this wage is high, a larger share of adjustment to L_{CSI} falls on formal employment relative to countries where it is low and where more shifting can in principle take place.

The available studies for Latin America find large variance in the magnitude of shifting. Gruber (1997) finds almost full shifting in Chile. Cruces et al. (2010) find a milder response in the case of Argentina. Kugler and Kugler (2019) find that in Colombia, where the minimum wage is quite relevant, only 20 percent of contributions are shifted back to wages. Antón et al. (2012) calculate that in Mexico, where the minimum wage is less relevant, about two-thirds of contributions are paid by workers. Prada et al. (2015) show that women's wages fell between 9 and 20 percent when Chile introduced a mandate to provide childcare in firms with 20 or more female workers. See Pagés (2017) for a review.

Incidence analysis of L_{CSI} contributions usually focus on formal firms and workers. But labor market adjustments in response to these contributions also impact informal workers. When formal employment contracts, the supply of labor to the informal sector increases, depressing informal wages (Levy 2008). *De facto*, informal workers also pay for L_{CSI} programs, even though they receive no benefits. For them, contributions to social insurance result in a tax on their earnings. This effect can be substantial. Antón et al. (2012) estimate that in Mexico, L_{CSI} contributions reduce informal wages by 10 percent. And this effect further dilutes their efficacy to lower G(**Y**). How can these contributions lower G(**Y**) when low-income informal workers not only do not benefit at all, but are indirectly hurt by them?

The idea that firms' contributions to L_{CSI} reduce G(**Y**) is deeply engrained in Latin America. This idea was born in Germany at the end of the 19th Century, under very different conditions. The idea is central to the (L_{CSI} + L_{JS} + L_{MW}) regime which has been the cornerstone of social protection in the region for over 80 years. But the fact that the idea is long dated does not imply that it is right. We return to this issue in chapter four.

3.4.2. Contributory social insurance and inequality: cross-subsidies

We now discuss the second channel through which L_{CSI} programs could lower G(**Y**): cross-subsidies from high- to low-income workers. The first point to note is that these subsidies occur only among formal workers; by design, low-wage informal workers cannot benefit. The second point is that the magnitude and, in some cases, direction of cross subsidization, depends on the specifics of the various programs, particularly for health and retirement pensions.

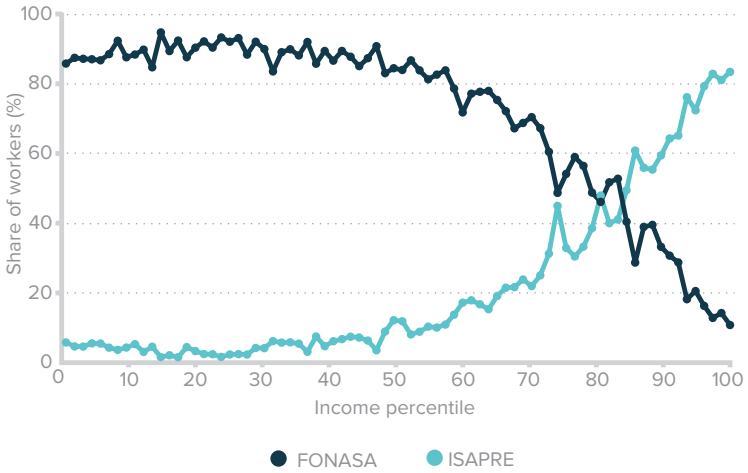
Consider health programs first. Cross-subsidies can be large when contribution rates and the quality of services is the same for all. In absolute terms, high-wage workers contribute more than low-wage ones, but they all benefit equally. Under these conditions, these programs serve to both pool risks and lower G(**Y**), even if their effects are confined to formal workers; this is the case in Colombia and Ecuador.

But various circumstances may weaken cross-subsidies. The first one is that many high-income earners do not participate: lawyers, doctors and other professionals are excluded from the obligation to contribute because they are self-employed; and entrepreneurs who in most countries are also excluded from L_{CSI} laws. A second one is that the contribution rate for low-wage workers may be higher than for high-wage ones, as occurs in Mexico, where those earning one minimum wage contribute 19 percent of their salary, while those

earning 10 contribute 5 percent. A third reason is that laws may cap contributions for higher-wage workers. Honduras provides an extreme example because contributions are capped at one minimum wage. Since workers must earn at least that amount to be formal, and since all contribute on that basis, there is risk pooling but no cross subsidization; *de facto*, redistribution is ruled out by law. And yet a fourth reason might be large co-payments for medicines and services that imply that contributions do not fully cover costs and absorb a larger share of the income of lower-wage workers, as in the Dominican Republic.

In other countries, like Argentina, Chile and Peru, cross-subsidies are weakened because regulations allow high-income workers to channel their contributions to separate funds. In Argentina, formal workers contribute 9 percent of their salary to health insurance, but those with high wages can channel their contributions to private providers, with labor unions acting as intermediaries. In Chile, workers contribute 7 percent of their salary for health insurance. They then have the option of diverting their contributions to private insurers known *Instituciones de Salud Previsional*, or Isapres, or to pool them in the public national health fund *Fondo Nacional de Salud*, or Fonasa. In the first case, they can complement them with additional resources to get higher quality services. In the second, they may get some subsidies depending on their income level but receive lower quality care. Graph 10 shows the share of workers that contribute to each option by income level. Clearly, high-income ones do not pool their contributions in Fonasa, and avoid cross-subsidizing low-income ones. The result is that health programs in Chile do little to redistribute income from high to low-wage workers, and that all workers do not receive services of the same quality.[27] A similar but less severe situation occurs in Peru, where workers contribute 9 percent of wages to health, but are only allowed to divert 2.25 percent to private insurers.

Graph 10: Health Programs in Chile



What about contributory retirement pension programs? If the contribution rate in the PAYG modality is actuarially fair, there is no redistribution involved. The same holds in the defined

[27] The effect is large since those at the top of the distribution have substantially higher wages. Of all workers, 14 percent channel their contributions to the Isapres, for a total of 0.98 percent of GDP; the contributions of the remaining 86 percent that go to Fonasa are worth 1.18 percent of GDP.

contribution modality, where workers' pensions are proportional to the amounts accumulated in their individual accounts. But this is rarely the case in Latin America. Five features imply that these programs may have strong redistributive impacts, although not always in the right direction: conditions to qualify for benefits, actuarially imbalanced contribution rates in PAYG modalities, separate regimes for certain groups like public sector workers, indexing rules and links between pensions and health programs.

As noted, many countries require workers to contribute for a minimum number of years to qualify for a pension. What happens when they do not qualify? Depending on their country's laws, they may lose all or part of their contributions, or get them back in one lump sum (sometimes with interest, sometimes without). Put differently, workers' pension savings are expropriated or heavily taxed. Even if they are not, longevity risks are fully transferred to workers, partly nullifying one of the purposes of these pensions.

However, as also noted--and more relevant for inequality--, spells of formality are not neutral across income levels; because low-wage workers spend less time in formality than high-wage ones, they are more likely to be *de facto* taxed and forced to bear the risks of longevity.

A few examples are illustrative. In Honduras, workers who fail to qualify get back less than a third of what they contributed (a tax of about 71 percent on accumulated contributions), while those who do qualify receive a pension that is nine times larger than their contributions, partly financed with the contributions of those who did not qualify. Because low-income workers are less likely to qualify, they end up subsidizing high-income workers who do, making the scheme a regressive one indeed. In Brazil, those who do not qualify also lose all their contributions, a large tax considering that contributions are 28 percent of the wage. This is likely a regressive tax considering that low-wage workers are more likely to be in that situation. In Ecuador, workers who fail to qualify—again, typically lower-income ones—lose all their contributions; in parallel, workers who do qualify—higher-income ones—are subsidized because the government covers 40 percent of the value of their pensions. The situation in Peru is mixed: workers who do not qualify in their PAYG modality lose all their contributions (a 100 percent tax used to partly fund the pensions of those who do qualify), but they get back all their resources in their defined contribution modality in one lump sum.

On the other hand, when workers do qualify, there can also be redistributive impacts, particularly in PAYG modalities when the contribution rate is below the actuarially fair value of benefits. In these cases, the government fills the gap, and those who get a pension receive a subsidy. In principle, these subsidies can be progressive or regressive, depending on the income level of those who get them. In practice they are often regressive, because higher-wage workers are more likely to qualify for pensions.

These impacts can be large, as exemplified by Colombia, where the government spends 3.4 percent of GDP subsidizing PAYG pensions, including special regimes for various groups, and where 70 percent of these subsidies are captured by households in the three upper deciles of the income distribution. They are also large in Ecuador, where subsidies are 1.6 percent of GDP and where higher-wage workers are also more likely to get a pension. They are also large in Mexico, where about 75 percent of the value of the pension is subsidized, resulting in subsidies of 2.5 percent of GDP, also concentrated in the upper deciles of the income distribution (Scott, 2014; Berstein et al., 2018).

A third factor that impacts the distributional impact of contributory pensions has to do with special regimes for some workers, often those in the public sector. Brazil is a case in point. There are two separate PAYG regimes, one for private sector workers and a separate one

for public sector workers. Both are subsidized—2.8 percent of GDP for the former and 2.9 percent for the latter—but the former has 10 times more retirees. Similar situations occur in Mexico, where there are also separate regimes for private and public sector workers (Scott, 2014), and in Argentina, where there are a multitude of special regimes (Rofman, 2021).

Yet a fourth factor relates to indexing rules. In some countries, pensions must at least equal the minimum wage, so that when the minimum wage increases in real terms, the same happens to minimum pensions. Since pensions above that threshold are indexed to inflation, the pension distribution narrows, reducing G(**Y**), while the average real value of pensions increases. This is relevant in countries like Colombia, where the accumulated increases in the price level and the nominal minimum wage over the last 15 years have been, respectively, 77 percent and 130 percent. In other countries, however, indexing rules can result in more complex outcomes, particularly if they change overtime. In Argentina, pensions were indexed to tax revenues until 2017, implying that when revenues increased faster than inflation, they would increase in real terms. Since prior contributions are not affected, subsidies from taxpayers to those with a pension would have to increase. Conversely, if revenues lagged behind inflation, pensioners would *de facto* be taxed. While it is difficult to quantify the impact on G(**Y**) in each case, the point here is that sometimes it would be positive and sometimes negative. These rules changed in 2018, and pensions were indexed to a weighted index of wages and prices. As of 2021, however, they are once again indexed to tax revenues. Thus, in this case, indexing rules have erratic impacts on inequality.

Finally, an insufficiently noted aspect of contributory programs is that access to health services is sometimes linked to pensions. This happens in Argentina, Ecuador, Honduras, Mexico and Peru, because contributory health programs operate under a PAYG design: the contributions of active workers finance health services for retired ones. However, coverage for retired workers is limited to those who qualify for a contributory pension, or at least to those who contributed for a minimum number of years. But because of formal-informal transitions, many workers do not qualify. When they work formally, they finance health services for retired workers, but when they retire, they will not have access to such services. In Mexico, for instance, workers have to contribute for 15 years to be entitled to contributory health services upon retirement, but half will not reach that threshold. Since lower-wage workers are more likely to fall into these circumstances given their shorter spells of formality, they end up cross-subsidizing contributory health services for higher-income retirees with a contributory pension, while they are left without a pension and receive lower quality health services from non-contributory programs. This is another subtle but no less relevant mechanism that negatively impacts G(**Y**).

In sum, six factors lead to the contradictory effects of contributory social insurance programs on income inequality in the region:

- not all firm owners have higher incomes than all workers;
- firm contributions are partly shifted back to formal workers;
- informal workers are excluded from contributory benefits, but may indirectly pay for them in the form of lower wages;
- cross-subsidies from high- to low-income formal workers may be diluted or nullified;
- differences in contribution densities between high- and low-income workers mean that the latter are less likely to qualify for retirement pensions and lose all or part of their

contributions while also being denied access to contributory health services; and,

- public subsidies for pensions are sometimes disproportionately concentrated on higher-income workers, or public sector workers through special regimes.

The redistributive inefficacy of contributory social insurance programs in Latin America is a major problem. In other regions of the world, these programs, aside from providing insurance, play a central role in mitigating income inequalities. But in Latin America, their atrophy as a mechanism to provide insurance goes hand-in-hand with their atrophy as a mechanism to reduce G(**Y**).

3.4.3. Non-contributory social insurance and inequality

L_{NCSI} programs are more likely to reduce G(**Y**) because they are financed from general revenues and because, on average, informal workers have lower incomes than formal ones. Households of all income levels contribute through general taxation, but benefits are concentrated in lower-income ones; cross-subsidies operate in the desired direction. Moreover, cross-subsidies involve all households, including those that derive income from land or financial assets, and are thus more powerful than those that occur through L_{CSI} programs (when they occur).

Non-contributory programs are concentrated in health and retirement pensions. All countries have programs for health. In some cases, they are the most redistributive social protection program as judged by the resources spent on them: 2 percent of GDP in Colombia, 2.5 percent in Ecuador, 1.4 percent in Honduras and 2.4 percent in Peru.

That said, their impact on G(**Y**) is weakened partly because, as exemplified in Graph 6 for Ecuador and Peru and Graph 8 for Mexico, there is significant overlap in the earnings distributions of formal and informal workers, meaning that some subsidies can be captured by high-income informal workers. In fact, this overlap creates a troublesome, insufficiently noticed feature: formal low-wage workers are required to contribute for their health benefits, while informal higher-wage workers obtain them for free.

In this context, Brazil merits special attention because it is the only country where there is no distinction between contributory and non-contributory health programs. Brazil's L_{CSI} excludes contributions for health, and the country has a unified national public health program (*Sistema Único de Saúde*, or SUS) funded from general revenues. All workers have access to services of the same quality regardless of their labor status, and this case the distinction between contributory and non-contributory programs is meaningless (but not in pensions). The SUS is best thought of as a universal in-kind transfer program. Brazil spends 3.9 percent of its GDP on the program, and it is difficult to think of money better spent: universal coverage with equal quality regardless of income or the vagaries of the labor market.

Non-contributory pension programs also tend to reduce G(**Y**) for the same reasons as non-contributory health programs, although impacts are smaller because fewer resources are involved. They take many forms. The simplest scheme is a flat pension for all elderly, as in Bolivia and, more recently, Mexico. The amounts spent differ—1.1 percent of GDP in Bolivia and 0.6 percent in Mexico—but regardless, the scheme is highly redistributive in both countries. In fact, in these two cases, these pensions are best thought of as an in-cash universal transfer program. Aside from reducing G(**Y**), they contribute to a reduction in old-age poverty and help the elderly manage the risks of longevity. Further, they have the great advantage of

being wholly independent of previous participation in the labor market, avoiding any impact on the contribution decisions of currently active workers (as discussed in the next section).

In other countries, however, non-contributory pensions are limited either only to poor individuals (so that one can also think of them as targeted poverty programs), or to workers that did not get a contributory pension. Because of that, and because in general the pension is small, the resulting redistribution is not very significant, with the added complication that the conditions imposed can affect behavior. Colombia, for example, restricts its non-contributory pension to poor households and sets it at one-tenth of a minimum wage. As a result, it spends only 0.1 percent of GDP on this program. Brazil sets it at one minimum wage and spends more, 0.4 percent of GDP. But other countries spend substantially less: 0.11 percent in Peru, 0.009 percent in the Dominican Republic and 0.0007 percent in Honduras. These figures compare very unfavorably with subsidies to contributory pensions, which, as discussed, tend to favor higher-income households.

Chile's pensions merit special attention. They are under a defined contribution modality, but the government provides subsidies to ensure that all the elderly in households up to the sixth decile of the income distribution get a pension at least equal to a Basic Solidarity Pension (one-third the minimum wage), even if some never contributed to their individual account. Those who do contribute receive subsidies if the pension obtained with their own resources is below a threshold called the Maximum Pension with Solidarity Support (which is almost equal to one minimum wage). Subsidies decline as workers' pension savings increase but insure that, at the margin, workers are always better off contributing more. Further, workers never lose their contributions because there is no minimum period required to qualify for a pension. The scheme is perhaps best thought of as a semi-contributive one. But regardless of the label, the scheme reduces G(**Y**) because, even if all workers had the same spells of formality during their lifetimes, lower-wage ones would accumulate less than higher-wage ones, and thus receive larger subsidies; the fact that low-wage workers have lower contribution densities increases the redistributive impact. That said, the parameters of the scheme are not too generous, and altogether the government allocates 0.7 percent of GDP to subsidize 1.4 million retirees, compared with 1.35 percent to subsidize 773,000 retirees under the old PAYG modality. The latter group are by and large concentrated in the higher deciles of the income distribution (although the parameters have recently changed, and spending will increase).

3.4.4. Summing up on social insurance

Tracing the impact of the combination of contributory and non-contributory programs on G(**Y**) is complex because some firm owners can have lower incomes than some workers; because the same worker can sometimes pay for benefits through a contributory program and sometimes receive free benefits through a non-contributory one; because workers with identical observable characteristics (age, education) and almost the same accumulated years of formality can end up with very different outcomes, with one getting a subsidized pension and one losing part or all of their contributions; because workers with similar careers can get health services of different quality when they retire; because there can be intertemporal effects that impact the real value of pensions; and so on. These programs have embedded in them an implicit system of taxes and subsidies that redistribute income between workers of different cohorts, between workers and firms, and between workers and other households. These taxes and subsidies may change over time, are not neutral across income levels, and do not necessarily favor lower-wage workers. Unfortunately, the data required to identify and

measure them are often not available, even in countries with reasonable household surveys, in part because these surveys seldom have long enough panels to measure dynamic effects (for instance, workers who contribute but fail to receive contributory pensions or health services).

Further, these programs impact the distributions of firms' profits and workers' wages as firms try to shift L_{CSI} contributions back to workers, and as formal and informal wages adjust to the associated changes in the composition of employment. Thus, an assessment of the programs' impact on G(**Y**) requires not only tracing the implicit system of taxes and subsidies alluded to above, but also the adjustments in the overall income distribution, particularly between firm profits and workers' wages, and in the formal-informal composition of employment.

In this context, a region-wide assessment of the impact of social insurance on inequality is difficult, all the more so because programs that superficially look similar across countries can be very different once indexing rules and conditions of access and labor market dynamics are taken into account. The discussion above therefore should be read less as an overall assessment, and more as a discussion of the wide range of outcomes that are observed in Latin America given its dual social insurance architecture.

That said, on the whole, non-contributory programs tend to be more redistributive than contributory ones, even if their quantitative impact is small in some cases because the budgetary resources involved are not large. In turn, and again on the whole, health programs tend to be more redistributive than pension programs, because contributory programs usually (but not always) involve cross-subsidies from high- to low-wage workers, and because the budgetary resources allocated to non-contributory health programs are larger than those allocated to non-contributory pensions. Within pension programs, contributory programs are the most problematic, in many cases redistributing income in the direction opposite to what is desired, a situation aggravated when these programs demand subsidies from the government that exceed those allocated to non-contributory pensions, sometimes by a significant margin.

3.4.5. Other taxes and transfers and inequality

Governments establish other policies to reduce G(**Y**) beyond social insurance and targeted poverty programs. Personal income taxes stand out because if they are broad-based and the schedule is progressive, they can extract large resources from high-income households to fund programs for lower-income ones. With variations, the region's personal income tax schedules are all progressive. However, as documented in chapter four, a mix of loopholes and evasion substantially erodes the redistributive power of this tax. In this context, and in a context where social insurance programs have weak, ambiguous or even negative impacts on G(**Y**), countries implement many more policies to temper income inequalities. Three merit attention:

- programs like school breakfasts or meals, distribution of food or other staples, vouchers for childcare and so on, some universal and some targeted by income level;
- exemptions to the VAT and other consumption taxes, and generalized subsidies for gasoline, electricity and the like; and,
- special tax regimes for small firms and low-income self-employed workers.

These policies, overlapping with contributory and non-contributory programs, minimum wages, regulations on job stability and targeted income transfers in the form of CCTs, make the architecture of social protection in Latin America very complex and, in the end, less effective.

The standard mechanism to capture the impact of taxes and transfers on G(**Y**) is to distinguish between market income Y_i, as depicted in the last line of Graph 4, and disposable income ($Y_i - t_i$) where t_i is the net result of taxes paid and transfers received by the i^{th} individual. In general, t_i considers personal income taxes, contributions to and benefits from contributory programs, benefits from non-contributory programs, consumption taxes, targeted transfers from CCTs and the like, universal transfers (sometimes including education) and subsidies for energy. Importantly, t_i only captures the impact of social protection policies that can be directly assigned to individuals or households. Corporate taxes and special tax regimes for micro and small firms are usually excluded, and so are minimum wages.[28]

From the point of view of income inequality, one would like t_i to be positive for high-income individuals, so that their disposable income is lower than their market income, and negative for low-income individuals. More precisely, letting $\mathbf{t} = (t_1, t_2, ..., t_n)$, taxes (including contributions to social insurance) and transfers reduce inequality if:

$$(6)\ G(\mathbf{Y}) > G(\mathbf{Y} - \mathbf{t})$$

While results of relation (6) depend on the specific taxes and transfers considered in **t**, the available evidence suggests that the difference between G(**Y**) and G(**Y** – **t**) is smaller in Latin America than in the OECD. For the average OECD country, G(**Y**) equals 0.47 and G(**Y** – **t**), 0.30; the corresponding averages for Latin America are 0.51 and 0.49. In the first case, the difference is 0.17; in the second, 0.02.[29]

Moreover, although there have been variations over time, income inequality in Latin America has remained high. Lustig (2020) shows that for the region as a whole, the Gini coefficient was 0.474 in 1990 and 0.455 in 2017; see also IDB (2020a).[30] The reasons behind the persistence of high inequality vary from country to country and it is not possible to provide a general explanation. In most, part of the inequality of G(**Y**) derives from issues of access or quality of education. In some, competition and other policies to reduce rents at the top have not been as powerful as needed. In many countries, personal income taxes have weaknesses that limit their impact on G(**Y** – **t**). Further, unlike other regions, in the context of the formal-informal dichotomy, L_{CSI} programs have much more limited effects, and in some cases those effects are the opposite of what was intended. These limitations have only been partly overcome by non-contributory and targeted poverty programs and a myriad of tax exemptions and special tax regimes, accompanied in some cases by generalized subsidies for energy or lower VAT rates for certain goods.

---

[28] That said, for the reasons discussed, imputing the costs and benefits of contributory pensions is complex because household surveys usually do not record if a worker receives a lump sum payment at time of retirement rather than a flow of payments during retirement, nor the implicit taxes when workers fail to get a pension and lose all or part of their contributions; see also (Grushka, 2019). There are also challenges to impute costs and benefits of contributory health programs when they operate as a PAYG and workers contribute part of their working life, but do not qualify for services when they retire, or when they can deviate their contributions, as in Chile.

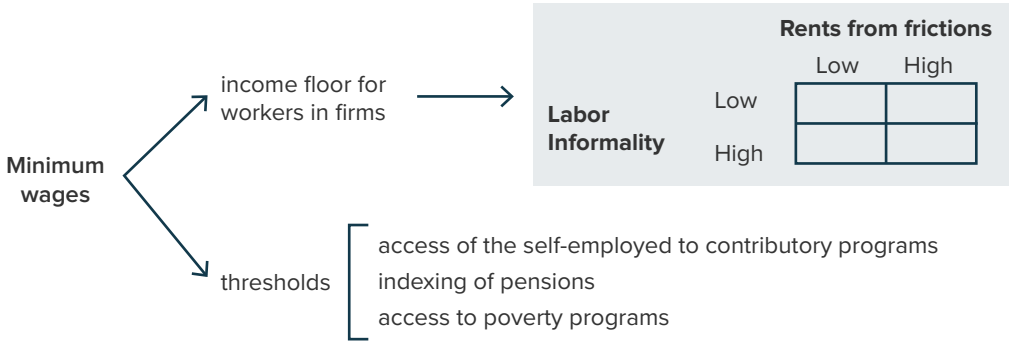
[29] Calculations for the OECD are from the standardized OECD database (Causa & Hermansen, 2019; OECD et al., 2020) and for the region from the CEQ Project at Tulane University (CEQ, 2020). The OECD calculations exclude from t indirect taxes and subsidies via provision of health or education. To make the numbers as comparable as possible, we compare the Gini coefficient of "market income" and "disposable income" of the CEQ project, so that in both cases **t** includes contributions to contributory programs, personal income taxes, and cash and in-kind transfers from non-contributory and targeted poverty programs. Country by country comparisons are also presented in IDB (2020a).

[30] Lustig also compares the Gini coefficient in 1990, 2000, 2010 and 2018 between Latin America and other regions of the world and shows that Latin America's was the highest each year.

3.4.6. Minimum wages and inequality: income floors for workers

Minimum wages in Latin America have two separate roles, as depicted in Graph 11: they affect the earnings of those actively working, and they are thresholds to determine access to some social protection programs.

Graph 11: Two Roles for Minimum Wages



Source: authors' elaboration.

Begin with the first role. When labor markets are perfectly competitive, minimum wages have contradictory impacts on G(**Y**): they raise the wages of low-income dependent workers that are formally employed, but also result in higher unemployment. The wage distribution narrows, lowering G(**Y**), but some workers are left without their salary, increasing G(**Y**). The results are therefore ambiguous and depend very much on the level at which the minimum wage is set, and the number of low-wage workers that benefit versus those displaced into unemployment.

Outcomes can be different when labor markets are not perfectly competitive because of various factors usually gathered under the label of frictions. In competitive labor markets, firms find workers rapidly and at no cost, and workers find jobs equally rapidly and also at no cost; in this context, wages are only determined by the productivity of workers. However, if finding workers and finding jobs is costly, the process of wage determination changes. Firms may have some power over workers because they know that if workers do not accept their wage offer, they have to search for another job, which may not be immediately available, and forgo some income; as a result, they can pay workers less than the value of their marginal product and enjoy some rents. On the other hand, workers may also have some power because they know that if firms do not fill their vacancies immediately, they will forgo some profits; as a result, workers can be paid more than the value of their marginal product and enjoy some rents. In a context where it is costly for firms and workers to search for better options, wages result from a mix of the productivity of workers and the relative bargaining power of the parties. Importantly, note that the source of rents is the search process itself: firms and workers know that it is costly for both to forgo an agreement because neither will find an alternative immediately.[31]

[31] As a result, firms could have power in the labor market even if product markets are competitive. On the other hand, firms could have power in the product market but not in the labor market if search costs are very low and there are no rents to extract from bargaining with workers. What matters for minimum wages is what happens in the labor market, not the product market. That said, if firms have power in the product market and face higher minimum wages, they can pass on part of the increase to consumers through higher prices, and the impact on G(**Y**) would be more difficult to trace.

In a search-bargaining context where frictions create rents, the minimum wage tilts the bargaining process in favor of workers. In particular, a binding minimum wage changes the distribution of rents in favor of low-wage workers. Moreover, even higher-wage workers benefit, as their bargaining position is also strengthened— a result known as the lighthouse effect, where the minimum wage becomes a reference point for all wage negotiations. This result is well established in developed countries, particularly the United States (Card & Krueger, 2015; Flinn, 2011).

Importantly, the fact that there are rents does not imply that a binding minimum wage has no impact on employment, although it does imply that it is less than if the market was perfectly competitive. Put differently, the presence of rents serves to cushion the effect of higher wages on unemployment, but in principle not to eliminate it. Neumark and Shirley (2021) review the evidence for the United States and conclude that it points to small but negative employment effects from minimum wages. On the other hand, Hoynes and Patel (2018) find that the minimum wage can stop large employers from capturing the benefits of the Earned Income Tax Credit designed to help low-income workers. However, a review of papers from other developed countries has found no effects; see Dube (2019). As expected, much depends on the value of the minimum wage and the institutional context.

What happens in Latin America? Flabbi (2021) develops a search-bargaining model that incorporates two important features of the region's labor markets: imperfect enforcement of labor regulations that generate informal salaried employment, and self-employed informal workers. He finds that introducing a binding minimum wage increases the wages of formal workers on the low end of the formal wage distribution. Further, wages of all formal workers increase because they all benefit from the increased bargaining power of low-wage workers (although proportionately less for those receiving higher wages). In addition, he finds that this effect extends to the informal wage distribution, although impacts are quantitatively smaller. In parallel, he finds that unemployment increases, although less than if the market was competitive. These findings suggest that the lighthouse effect also holds in settings with an informal sector, and that the presence of rent-creating frictions is the critical feature behind it.

That said, informality matters because:

- the magnitude of the lighthouse effect is inversely proportional to the size of informal employment (in the limit disappearing if there is no formal employment);
- informal employment includes self-employed workers and others who do not bargain with firms, and those earnings are not changed by a minimum wage;
- in many informal firms, the incomes of the owner and its workers are not that different, so higher wages imply little redistribution; and,
- informal firms may have an ownership structure where owners and workers are relatives, implying that workers are not paid wages (rather, the surplus of the firm is shared between participants), and that any redistribution occurs within the household.

What are the implications for inequality? Clearly, a binding minimum wage narrows the wage distribution, and this lowers G(**Y**). On the other hand, it widens the differences in earnings between the self-employed and those with a salaried job, increasing G(**Y**). Finally, the associated redistribution of rents from firm owners to salaried workers is not always equivalent to a redistribution from high- to low-income households, as some firm owners have lower incomes than some workers, particularly formal ones (see Graph 8). In this context, the impact

of the minimum wage on G(**Y**) is ambiguous and will depend on the size and composition of informal employment, the size and ownership structure of firms and, of course, the rents available for redistribution given the relevance of search costs and other frictions.

The matrix in Graph 11 captures four stylized situations. In the upper right corner are countries where search costs and other frictions result in large rents, and where informal employment is low. In these countries, introducing a binding minimum wage would be more likely to lower G(**Y**); there would mostly be gains for formal workers and a few informal ones, and these gains would by and large come from larger formal firms likely owned by individuals at the high end of the income distribution; moreover, because rents are high, there would be a small impact on unemployment. In the lower left corner are countries with the opposite combination, where it is more likely that the minimum wage would increase G(**Y**) because most low-wage workers are informally employed, because most firms are small, informal and likely owned by individuals who are not in the upper deciles of the income distribution, and because there are few rents to redistribute, implying a larger impact on unemployment.

These are, of course, stylized situations and it is not easy to map individual countries in the region into the four quadrants of the matrix because the level and composition of informal employment between salaried and self-employment varies, and because it is difficult to measure the size of rent-induced frictions.[32] It is illustrative in this context to consider Table 6, constructed from Tables 1 and 2, which lists the overall informality rate, the share of the labor force that is self-employed, and the share employed by firms with five or fewer workers (except Honduras, where it refers to firms with up to 10).

Table 6: Informal Employment, Self-employment and Employment in Small Firms

	Informal employment	Self-employment	Employment in small firms
Argentina	36	27	24
Brazil	37	34	15
Chile	35	23	15
Colombia	63	34	25
Dominican Republic	59	52	9
Ecuador	68	35	34
Honduras	82	40	46
Mexico	57	21	30
Peru	77	46	18

Unfortunately, the data are insufficient to distinguish formal from informal firms in all countries, although by and large small ones are mostly informal everywhere (since the difference between overall informal employment and informal self-employment occurs in firms). The table highlights large differences in informality rates (35 percent in Chile versus 82 percent in Honduras) and in the composition of informal employment. On the other hand, a significant number of workers

[32] Rents depend on various parameters of the bargaining environment: rates at which unemployed workers receive offers from firms, rates at which jobs are destroyed, the utility on searching from unemployment and so on (see Flabbi, 2021). However, it is not easy to translate these parameters into a money metric of rents.

is employed in firms with 50 or more workers: 37 percent in Chile, 23 percent in Colombia and 21 percent in Mexico; or with 100 or more workers: 10 percent in Ecuador and 16 percent in the Dominican Republic. The firms that hire these workers are in all likelihood formal and the income of their owners is substantially higher than that of their workers.

The point here is that firms in Latin America are very heterogenous in terms of formality status, size and ownership structure and that a large number of low-income workers are self-employed or employed by informal firms—a very different context from the United States or other countries in the OECD, which is where the majority of studies on the impact of the minimum wage in a bargaining context have been carried out.

Firm heterogeneity is problematic because the minimum wage impacts all firms in a country. Some large ones may pay their workers below the value of their marginal product, but many more probably do not.[33] As a result, minimum wages in Latin America can redistribute rents from a few firms to a few low-wage formal workers and, at the same time, depress the earnings of lower-income informal workers, or increase unemployment, with offsetting impacts on G(**Y**). Much depends on country circumstances and parameter values.

Much also depends on the level at which the minimum wage is set, another dimension on which countries in the region differ. In Honduras, it is set at the 60th percentile of the wage distribution; in Ecuador, Colombia and Peru at the 50th; in Brazil, at the 35th; in Mexico, at the 25th; in Argentina, at the 19th; and in Chile, at the 15th. Given these differences, and differences in informality rates, it is more likely, for instance, that a binding minimum wage increases G(**Y**) in Honduras, while in Mexico it does the opposite, although in all unemployment and informal employment is higher as a result of it.

All this helps to explain why the evidence of the impact on the minimum wage on G(**Y**) in Latin America is mixed. For instance, Bosch and Manacorda (2010) find that the fall in the real minimum wage raised G(**Y**) in Mexico. On the other hand, Ham (2018) finds that higher minimum wages resulted in lower informal wages, more non-compliance by firms and an increase in poverty among informal workers in Honduras. Sometimes results also vary within the same country, depending on growth and overall macroeconomic conditions. In Brazil, for instance, Ferreira et al. (2021) estimate that minimum wage increases were associated with higher inequality between 1995 and 2002, but lower inequality between 2002 and 2012.

3.4.7. Minimum wages and inequality: thresholds

The lower segment of Graph 11 calls attention to a less discussed role of minimum wages in Latin America that also impacts G(**Y**): its use as a reference point for various social protection programs. The graph suggests three possible channels. The first one was already discussed in the section on poverty. Clearly, a high minimum wage makes it more difficult for the low-income self-employed to participate in contributory programs, and this would tend to increase G(**Y**).

---

[33] Some countries, like the Dominican Republic and Honduras, try to deal with this heterogeneity by setting higher minimum wages for larger firms, assuming they have more power in the labor market than smaller ones. However, as discussed in the next two sections, these policies deter firm growth and can hurt **TFP**. Countries also set minimum wages by region rather than firm size, and if regions are chosen carefully, this may be more effective and have lower **TFP** costs. Mexico, for instance, sets a higher minimum wage in the north versus the rest of the country.

The second channel depends on which pensions are indexed to the minimum wage. In the case of contributory pensions, a higher real minimum wage can increase G(**Y**) if pensions are subsidized by the government and are mostly concentrated in higher-wage workers. However, in some countries, like Brazil, the minimum wage is also the floor for non-contributory pensions, and thus raising it reduces G(**Y**) because retirees who were informally employed before retirement also benefit (this also lowers the incentives of current informal workers to contribute). On the other hand, some countries do not allow contributory pensions below a minimum wage, and raising it makes it more difficult for worker to obtain a pension.[34]

The third channel reflects a variety of circumstances that can impact G(**Y**) in opposite directions. For instance, in Colombia, self-employed workers that earn more than one minimum wage are required to contribute to L_{CSI} programs; those earning below are not and qualify for the country's non-contributory health program (*Regimen Subsidiado de Salud*). Thus, on the one hand, raising the minimum wage makes it more difficult for the self-employed to contribute; but, on the other, it increases the number of people who can access fully subsidized non-contributory benefits. Other examples could be cited.

To sum up: minimum wages in Latin America have complex effects on income inequality. To assess them one needs to consider their effects not only on dependent workers but also on other workers whose earnings may be affected by them, particularly the self-employed. Further, one also needs to consider their effects on other groups like the poor or retirees, whose incomes or access to social protection benefits may also depend on minimum wages. Two key points to be kept in mind are:

- even if, as in developed countries, frictions in the labor market mean that minimum wages can redistribute rents from firms to workers, the impact of that redistribution on inequality will be different because of differences in the size, formality status and ownership structure of firms; and,
- unlike in developed countries, the impact of minimum wages in Latin America extends beyond its direct effects in the labor market when it is used as a threshold for other policies.

As a result, one needs to distinguish the impact of minimum wages on: (i) the distribution of wages of formal and informal dependent workers; (ii) the distribution of earnings among all workers, including the self-employed and the unemployed; (iii) the distribution of profits and labor earnings; (iv) and the distribution of income over all individuals, including those that are retired or depend on programs where access is linked to the minimum wage. Studies usually capture the effects of minimum wages on (i) and (ii), but it is important to also consider their impact on (iii) and (iv) to obtain a broader perspective of their efficacy to reduce inequality.

One last remark: discussions of the minimum wage in Latin America are at times complex. On one hand, it affects many groups beyond workers. On the other, even if the focus is only on workers, the minimum wage is set with two objectives that are not always clearly separated.

---

[34] Some countries get around that requirement creating a parallel system of savings for retirement to give low-wage workers income during retirement proportional to the amount saved—but, legally speaking, not a pension. This is the case in Colombia, which has a system called "Periodic Economic Benefits" (Beneficios Económicos Periódicos, or BEPS). The result is that Colombia has co-existing three contributory pension systems: PAYG, defined contribution in individual accounts, and BEPS.

One, as in the discussion above, is to strengthen the bargaining power of workers *vis-à-vis* firms; and the second is to provide salaried workers with a basic living standard. The level of the minimum wage associated with each objective is different: in the first case it depends on the extent of frictions in the labor market and the composition of employment, and in the second, on the cost of an exogenously determined basket of commodities. This may help explain why, to take the extremes, the minimum wage in Chile is set at the 16th percentile of the wage distribution, and in Honduras at the 60th.[35]

Looking forward, it would be better if the two roles of the minimum wage were separated, and objectives more clearly identified. In our view, minimum wages need to recognize more fully the characteristics of countries' labor markets. This does not mean that all salaried workers should not attain a basic living standard—far from it; but it does mean that the policies to achieve this should not necessarily rest on the minimum wage, and should extend to all workers, not only salaried ones. We return to this discussion in the next chapter.

3.5. Social Protection Policies and Productivity

We now turn to the impact of social protection policies on productivity, ignoring considerations of insurance, poverty or inequality. Relation (5) underlines that resource allocation **R** and aggregate productivity **TFP** depend on all policies in $\mathbf{E}\{.\}$. This section focuses on those related to social protection, without underestimating the importance of other ones. We discuss two channels through which they impact **R** and **TFP**:

- the formal-informal composition of the economy; and,
- the use of countries' human capital, the incentives to invest in it, and the opportunities to enhance it through on-the-job learning and training.

3.5.1. The tax on formality and the subsidy to informality

Begin with the first channel. Table 7 identifies the costs to firms and the benefits to workers of three contractual modalities. The first occurs when they comply with (L_{CSI} + L_{JS} + L_{MW}) laws, so that both are formal, and the second when firms evade these laws, so that the firm is illegal and informal and the worker informal. The third is relevant to countries with exclusions in these laws: individuals employed in their own one-person firm; firms and workers without relations of dependency and subordination; or legal ambiguities coupled with tacitly accepted social norms that allow firms to ignore these laws without risk of being punished. Firms and workers in the first case constitute the formal sector, and in the second and third, the informal sector—partly legal, partly illegal and partly ambiguous.

[35] Of course, the minimum wage can also result from negotiations between representatives of firms and unions, which in Latin America are usually mediated by the government since the resulting wage needs to be validated by a law or presidential decree. But even in these cases, it would be good if policymakers had clarity with regard to the objectives and trade-offs involved.

Table 7: Contractual Modalities Between Firms and Workers

	Cost to firms	Benefits to workers	Implicit tax (+)/subsidy (-)
Formal	$w_f + (L_{CSI} + L_{JS})$; $w_f \geq w^{min}$	$w_f + v_{CSI}.L_{CSI} + v_{JS}.L_{JS} + T_{TRN}(?)$	$[(1 - v_{CSI}).L_{CSI} + (1-v_{JS}).L_{JS}] - T_{TRN}(?)$
Informal & illegal	$w_i + L_{ENF}.F$	$w_i + v_{NCSI}.L_{NCSI} + T_{TRN}(?)$	$L_{ENF}.F - [v_{NCSI}.L_{NCSI} + T_{TRN}(?)]$
Informal & legal	w_i	$w_i + v_{NCSI}.L_{NCSI} + T_{TRN}(?)$	$- [v_{NCSI}.L_{NCSI} + T_{TRN}(?)]$

Source: adapted from Levy (2018).

To avoid introducing new notation, for the purposes of Table 7 we interpret L_{CSI} as the cost of contributory social insurance programs and L_{JS} as the (contingent) cost of job stability regulations.[36] We also interpret L_{ENF}—the functioning of the institutions enforcing ($L_{CSI} + L_{JS} + L_{MW}$) laws—as the probability that those who violate them are detected and forced to pay fines F.[37] Similarly, here L_{NCSI} and T_{TRN} are the monetary costs of non-contributory programs and transfers, with the (?) added to T_{TRN} to capture the fact that some transfers may not apply in all cases because they may be conditioned on some status (informal/poor). In parallel, three parameters—all measured as a share of costs—capture how workers value the benefits derived from their labor status: v_{CSI} from contributory programs, v_{JS} from job stability protections and v_{NCSI} from non-contributory programs. Finally, in more usual notation, w_f and w_i stand for the formal and informal wage respectively, and w^{min} for the minimum wage.

Parameters v_{CSI} and v_{JS} determine the value that workers attach to the non-monetary benefits of the ($L_{CSI} + L_{JS} + L_{MW}$) regime. Here we treat them separately, but for dependent workers they should really be thought of as one since, as mentioned in chapter two, firms that hire workers under relations of dependency and subordination must comply with L_{CSI} and L_{JS} at the same time, and therefore the contract between them needs to internalize their costs. Further, note that v_{CSI} refers to all programs included in L_{CSI} since they are bundled by law; workers cannot choose to contribute to some but not to others.[38]

With these observations, start with the first row, formality. Firms must pay workers at least the minimum wage, pay for contributory programs and comply with regulations on job stability. Firms will only hire workers formally if their marginal revenue product exceeds [$w^{min} + (L_{CSI} + L_{JS})$]. This may be an important barrier. If the TFP_i of the firm hiring the worker is low, the marginal revenue product of the worker will be low as well (relative to [$w^{min} + (L_{CSI} + L_{JS})$]). In

[36] L_{CSI} is usually expressed as a share of the wage, but to simplify the notation here we express it as a fixed amount. Laws set L_{JS} as a one-time payment contingent on firing, but here we express it as the equivalent flow payment that a firm needs to pay considering the probability that at some future date it may dismiss a worker; see Heckman and Pagés (2004). Alaimo et al. (2017) show that the expected average value of L_{JS} in the region is 8.4 percent of the wage, a considerable amount. That said, the formulation used here fails to capture other impacts of L_{JS}, like the short- and long-term composition of contracts. It also assumes that firms are risk neutral.

[37] So that when $L_{ENF} = 1$, laws are perfectly enforced. Fines F must be larger than the contributions evaded, that is, that $F > (L_{CSI} + L_{JS})$, so that under perfect enforcement firms do not evade since it is cheaper to hire legally.

[38] This point is at times missed when pensions are analyzed on their own. Even if they are actuarially fair in PAYG systems, or workers receive the full value of their contributions in defined contribution systems, workers can still undervalue contributory social insurance because they are also required to pay for health, daycare, training or other programs. Some pension reforms in the region have been motivated by the desire to bring contributions and benefits closer together. But these reforms cannot fully achieve this on their own given the bundled nature of L_{CSI}. Further, these reforms have been undermined by L_{NCSI} programs as workers could receive some benefits for free. Levy (2008) emphasizes the importance of the bundled nature of L_{CSI}.

this case, the firm will only hire the worker formally if it can shift back to the worker a share of ($L_{CSI} + L_{JS}$) in the form of lower wages. But if w^{min} impedes that, the firm will not hire the worker, or will hire them informally.[39]

Importantly, a TFP_i above a certain threshold is necessary but not sufficient for firms to hire formally. Independently of w^{min}, laws on L_{JS} may also deter formal hiring if negative output shocks or labor-saving technical change are not considered just causes for dismissal, as is the case in most countries in Latin America. To minimize future firing costs, which can be extremely high, firms limit formal employment and use a combination of strategies to deal with short-run fluctuations: subcontracting to other firms, rotating workers through temporary contracts exempt from L_{JS} laws, or pretending that there is no relation of dependency and subordination. Whatever the trick, formal employment is deterred, particularly long-term employment.

Assuming firms hire formally, workers receive wages and are entitled to contributory programs and job stability provisions. If they value both at their costs, so that $v_{CSI} = v_{JS} = 1$, then what firms pay for their workers is the same as what workers receive. But if v_{CSI} or $v_{JS} < 1$, the result is an implicit tax on formal employment, as the last column in Table 7 shows. This tax, very importantly, is not associated with personal income taxes. Rather, it results from workers' real-life experience with the institutions designed to help them when they lose their job, and provide them with health services, pensions and whatever other benefits are included in L_{CSI}. Put differently, this tax reflects the fact that, independently of the promises contained in the letter of L_{CSI} and L_{JS} laws, workers consider that benefits are less than costs.

The reasons why $v_{CSI} < 1$ vary across countries.[40] One may be low quality of health services. A second is that workers are forced to save for a pension that they may never get—losing all or part of their savings— or contribute to the health services of retired workers but not be entitled to these services when they retire. A third is because workers consider that the private firms that administer their individual retirement accounts charge high fees, and that they are better off saving in a bank. A fourth is because when a share of workers' wages is deposited in individual accounts for severance pay, the interest rates are paid are below market values. A fifth is because some programs in L_{CSI} may imply cross-subsidies among formal workers. And yet a sixth one is that other programs in L_{CSI} may imply cross-subsidies from all formal workers to the rest of the population.[41]

---

[39] Kugler and Kugler (2009) provide an example of the relevance of w^{min} in their analysis of Colombia's 1993 health and pension reform. An additional complication occurs in countries where labor regulations require firms to pay one (as in Mexico) or two (as in the Dominican Republic and Ecuador) additional months of salary for each full year of work. If w^{min} is not binding, firms adjust wages so that the only effect of the regulation is in the time-profile of payments; but if it is binding, this is not possible and *de facto* the minimum wage is 8 or 16 percent higher than what is actually stipulated.

[40] Bobba et al. (2017) estimate that in Mexico $v_{CSI} = 0.55$. Cuesta and Olivera (2014) estimate that the willingness to pay for L_{CSI} programs in Colombia is less than 50 percent below par, i.e., that $v_{CSI} < 0.5$. Other studies find increases in formal employment in response to increased workers' valuation of L_{CSI}; for instance, Bérgolo and Cruces (2014) in Uruguay. Similarly, Kugler et al. (2017) and Bernal et al. Pinzon (2017) find that formal employment increased in Colombia when the 2012 health reform lowered firms' contributions to L_{CSI}, consistent with $v_{CSI} < 1$ (otherwise wages would have increased by the full amount with no change in employment). Further evidence is provided by the fact that in many countries, workers legally excluded from the obligation to contribute to L_{CSI} programs can voluntarily contribute. The fact that hardly any do is a clear revelation of their preferences, i.e., that $v_{CSI} < 1$.

[41] Brazil's L_{CSI} includes contributions for education and agrarian development programs; Colombia's for early child development, education and housing; Mexico's for cultural and sports facilities; and Peru's for labor training. In all cases, these programs benefit all workers.

Similarly, many reasons explain why $v_{JS} < 1$. Workers may find that when firms fire them without just cause, the courts that administer job stability regulations are slow, or corrupt, delaying or diminishing the compensations that they are entitled to—as the findings of Kaplan et al. (2008) and Kaplan and Sadka (2011) show for Mexico. Or they may find that when firms go bankrupt (or simulate bankruptcy), it is impossible to collect those compensations.

Many others could be cited. But whatever the reasons, the result is that when v_{CSI} or $v_{JS} < 1$, formal employment is implicitly taxed. Regardless of the incidence of contributions, workers and firms jointly pay ($L_{CSI} + L_{JS}$) for non-wage benefits, but workers get only a fraction of that.

Consider next the second row, illegal informality. If firms cheat, their labor costs can be less than [$w^{min} + (L_{CSI} + L_{JS})$], even less than w^{min}. This means that firms with low TFP_is can survive; put differently, cheating allows firms with very inefficient technologies to exist. This is part of the explanation for why goods and services in Latin America are still produced with very old machines and almost obsolete capital goods, or why technologies that were abandoned long ago in developed countries still operate in the region.

When firms cheat, workers lose the protections of formality; instead, they can benefit from non-contributory programs L_{NCSI}. As before, workers may not fully value these programs, so $v_{NCSI} < 1$ (low-quality services, rationing and so on). But as long as $v_{NCSI} > 0$, informal employment is subsidized. Workers get benefits that neither they, nor the firms hiring them, pay for. Of course, whether firms cheat or not depends on the fines imposed if caught, F and on L_{ENF}, the probability of being caught.[42] But if L_{ENF} is small, firms will cheat, and illegal behavior will be subsidized because workers get free social benefits despite the fact that firms break the law.

Finally, turn to the last row, legal informality. In this case, w^{min} is irrelevant. Because there is no lower bound on what firms can pay their workers, some with very low TFP_i can legally exist. This is another reason that explains why very inefficient technologies are deployed in the region; their low labor costs allow them to exist.

One-person firms, or firms without dependent workers, are clearly subsidized by L_{NCSI} programs. The same occurs in firms where the relation of dependency is ambiguous, particularly if countries' tacitly accepted social norms *de facto* ensure that they are excluded from the enforcement of ($L_{CSI} + L_{JS} + L_{MW}$) laws; for them, $L_{ENF} = 0$. But regardless of whether firms and workers are in the second or third row of Table 7, the important result is that L_{NCSI} programs increase informal employment.[43]

In some countries, transfers from poverty programs in T_{TRN} are only conditioned on workers having an income below $\mathbf{Y}^{cutoff}$, but are independent of their formal or informal status. In this case, T_{TRN}(?) would be positive in all three rows of Table 7 for poor workers, and these programs would have no bearing on the formal-informal composition of employment. However, in other

[42] It depends also on v_{CSI} and v_{JS} because if they equaled unity, there would be no incentive to cheat. In fact, the existence of cheating is indirect evidence that workers undervalue the benefits of contributory programs and dismissal regulations because if $v_{CSI} = v_{JS} = 1$, there would be no incentive to evade them.

[43] Garganta and Gasparini (2015) find that in Argentina the extension of child allowances to informal workers resulted in lower transits from informality to formality; and Bosch and Guajardo (2012) find that the expansion of non-contributory pensions through the "Moratoria" program reduced formal employment among women. Camacho et al. (2014) find that the expansion of Colombia's non-contributory health program in the early 1990s increased informal employment. Bosch and Campos-Vázquez (2014) provide a similar finding for Mexico: the introduction of a non-contributory health program in the early 2000s reduced firms' compliance with L_{CSI} laws and formal employment; see also Bosch, Cobacho and Pagés (2014).

countries, like Ecuador and Peru, these programs are also conditioned on informal status. In this case, T_{TRN}(?) would be zero in the first row of Table 7, and positive in the other two. Clearly, if poor workers can only access some benefits if they are informally employed, informal employment among poor workers will be higher, and so will overall informal employment.

The message of Table 7 is both simple and very important. *Mutatis mutandis* across countries, social protection policies:

- create barriers to enter formality;
- contingent on entry, they implicitly tax it; and,
- subsidize informality.

3.5.2 Special regimes for the self-employed and micro and small firms

Many countries have special tax and social insurance contribution regimes for small firms or self-employed workers, with multiple objectives: reducing inequality by subsidizing low-income entrepreneurs, facilitating the incorporation of the self-employed into contributory programs and providing facilities for small firms to comply. These regimes can be considered special cases of L_{CSI} or income tax laws, although in some countries the border between them is blurred because tax and social insurance obligations are merged in a single payment. There is large variance across countries, and sometimes there are multiple regimes in the same country; see Kanbur and Keen (2014) for a general analysis.

Azuara, et al. (2019) provide an overview of these regimes in Latin America, but it is useful to give some examples. In Argentina, there are three regimes for the self-employed, depending on their income level. In two of them, for those with lower incomes, tax and contributions are merged in a single tax (*Monotributo*) that is highly subsidized relative to the associated social insurance benefits; but in the third, workers are heavily taxed since they pay contributions, and income and value added taxes. Separately, dependent workers, together with the firms that employ them, pay 42 percent of wages in L_{CSI} contributions in addition to income taxes. The result is that the gross earnings that a worker needs to be paid to receive the equivalent of one minimum wage of income net of taxes and contributions is 0.79 of the minimum wage in the most favorable regime (*Monotributo Social*), 1.10 in the second-most favorable (*Monotributo*), 1.72 as a dependent worker, and 1.95 as a self-employed under the third, heavily taxed regime (*Autonomo*).

Brazil has three special regimes. In the Simplified System of Contributions, self-employed workers can comply with their L_{CSI} obligations at substantially lower rates than dependent ones, as long as they do not sell to firms. If they want to sell to firms, they must register as a firm in the *Micro-empreendedor Individual* (MEI) regime and are allowed to employ an additional person (*de facto*, a two-person firm), but only if sales are below a given threshold. If sales exceed that threshold but are still below a second higher one, firms can qualify for a third regime called *Simples Nacional*, where payroll, income taxes and value added taxes are merged. After that second threshold, firms must pay L_{CSI} contributions and income taxes under the general regime. Altogether, there are seven alternatives with substantial variations in contribution rates and tax burdens. The result is that the productivity of a worker receiving the same net wage can vary up to 64 percent depending on whether they are hired as a dependent under the general regime (the most onerous), or the MEI (the least onerous).

In Peru, self-employed workers are not required to contribute to L_{CSI} programs, but firms face three regimes depending on annual sales: micro, up to US$180,000; small, up to US$2,000,000; and large, with substantial differences in the costs of L_{CSI} and L_{JS} programs, as shown in Table 8. Firms qualify for the first two regimes if they are below the sales thresholds and have fewer than 10 workers.

Table 8: Special Regimes for Firms in Peru

	Micro	Small	Large
Contributions to social insurance (L_{CSI}) (Share of wages, including additional mandated benefits)	18	36	50
Severance pay (L_{JS}) (Months of salary by tenure)			
1 year	0.3	0.7	1.5
2 years	1.7	3.3	7.5
3 years	3.0	4.0	12.0

Recall, however, that a self-employed worker can be a micro firm, particularly when those helping him are relatives without a relation of dependency; and recall as well that in Peru, 10 percent of the labor force is made up of non-remunerated family workers. Thus, in Peru there are really four social insurance regimes for firms, although only three are explicitly recognized, with the fourth one de facto exempt from all obligations derived from (L_{CSI} + L_{JS} + L_{MW}) laws. In addition, firms pay income taxes under four regimes, not shown in Table 8, that also depend on sales and number of workers (Cooper & Cuba, 2019).

Ecuador is also interesting. There are two special regimes (but really three considering that the self-employed are not required to contribute). In the *Régimen de Incorporación Fiscal Simplificado de Ecuador* (RISE), if sales are below US$60,000 a year, a firm can consolidate VAT and corporate taxes in a single payment that depends on the sector of activity and the value of sales. In addition, for every worker hired formally, it gets a 5 percent discount on the consolidated payment, cumulative for up to 10 workers, so that at that point the tax burden is halved. However, if the firm hires an additional worker, it no longer qualifies for any discounts. The result is that the cost of the eleventh worker is extremely high, as the firm's tax burden is doubled. On the other hand, if sales exceed US$60,000 but are below US$300,000, the firm no longer qualifies for RISE, but can qualify for the second regime, under which it pays corporate taxes at 2 percent of sales; but if sales exceed the second threshold, the firm pays 22 percent of profits under the general regime. In parallel, if firms have more than 10 workers, they have to set up stores where workers can buy basic products at cost; if they have more than 50, they also have to provide day care services for workers' children.

Honduras and the Dominican Republic provide yet another example: minimum wages that increase with firm size. In Honduras there are four tiers, with a 28 percent difference between the lowest tier (for the smallest firms) and the highest (for the largest ones). In the Dominican Republic there are three, with a difference of 59 percent between them.

Special regimes sometimes interact with other taxes, particularly the VAT. When firms are allowed to consolidate income and value added taxes in a single payment (sometimes together with contributions to social insurance), they may not be allowed to sell to firms that pay VAT in the general regime. The reason is that the VAT works through the credit method, where firms selling intermediate inputs issue invoices including the VAT, buying

firms deduct VAT paid on intermediate inputs so that consumers of final goods only pay tax on value added. When taxes and contributions are consolidated in a single payment, firms in the general regime cannot purchase intermediate inputs from firms in special regimes since the latter cannot issue invoices. The specifics vary from country to country, but the broad result is that the interphase of the VAT with the special regimes restricts firms in these regimes to sell only to final consumers, blocking backward linkages from larger firms in the general regimes.[44]

Table 9 provides a stylized summary. The point to highlight is that as a result of special regimes:

- workers may be paid below the minimum wage and dismissed freely;
- firms' labor costs, including the contingent costs of dismissal, are substantially below those they would face if sales or number of workers were above the threshold;
- firms face much lower corporate tax rates, which may be calculated over sales rather than profits, or even be a flat fee;
- firms may be able to consolidate social security contributions and income taxes (and sometimes value added taxes) into a single payment; and,
- firms may not be able to sell to other firms, but only final consumers.

Table 9: Special Regimes for Small Firms

	General Regime: sales or number of workers above a threshold	Special Regimes: sales or number of workers below a threshold
Relations of dependency and subordination with workers:	Yes.	Not relevant if truly self-employed. Ambiguous if firm involved, but *de facto* no due to social norms.
Subject to (L_{CSI} + L_{JS} + L_{MW}) laws:	Yes.	Truly self-employed: L_{CSI} with lower rates. Firm: L_{CSI} maybe with lower rates but L_{JS} and L_{MW} may not apply.
Labor costs:	$w + (L_{CSI} + L_{JS})$; $w \geq w^{min}$	$w + L_{CSI}$ with lower rates + maybe L_{JS} also with lower costs
Income taxes:	Normal rate on profits under corporate tax schedule. Withholding of workers' personal tax under standard schedule	Personal taxes at times consolidated with L_{CSI} contributions, or corporate taxes with a reduced rate on profits or rate based on sales or flat fee.
Consolidation of contributions and taxes in a single payment:	No.	Sometimes.
Selling to other firms	Yes.	Restricted sometimes.

Source: authors' elaboration.

[44] Preliminary results by Gerard et al. (2018) show that in Brazil trade links between firms occur mostly within firms in each regime. These results are in line with previous findings for Brazil showing that informal firms basically supply from informal firms (De Paula & Scheinkman, 2010). In Mexico the same issue is present, aggravated by the fact that, for social protection reasons, the VAT has three different rates (Levy, 2018).

From the perspective of **TFP**, these regimes have two worrisome effects:

- firms with very low TFP_is can legally exist as long as they are under the thresholds since their labor costs and tax burdens are substantially lower than larger, higher-productivity firms above the thresholds; and,
- if firms improve their productivity and their sales or labor force grow, placing them over the thresholds, their labor costs and tax burden increase, sometimes quite sharply, making growth potentially unprofitable.[45]

Of course, the importance of these effects depends on parameters: the size of the thresholds, the differences in tax rates between regimes, and so on. But, as exemplified by the Dominican Republic and Mexico, they can be very costly for **TFP**. In the Dominican Republic, there are 10 special regimes, depending on firms' size, location and sector. Azuero et al. (2019) use the country's administrative tax records to follow firms over a ten-year period, from 2007 to 2017, separating them between those in the general regime and those in special ones. They find that, on average, firms in special regimes have lower TFP_i than those in the general regime, but that they survive longer because their tax burden is substantially lower. They estimate that one peso of capital and labor invested in firms in the general regime is 43 percent more productive than the same peso invested in firms in the special regime.

Mexico has only one special regime for small firms, defined as those with annual sales below US$100,000. That threshold is sufficiently generous so that 93 percent of all firms captured in the economic census qualify, in which case they only pay 2 percent of sales in taxes, as opposed to 35 percent of profits under the general regime. Most firms in the special regime are very small, but because they number about 3.1 million, they jointly account for 52 percent of employment and 25 percent of the capital stock. The vast majority have very low TFP_is, but as in the Dominican Republic, they survive partly because of the special regime. Leal (2014) finds this regime reduces **TFP** in Mexico by between 19 and 34 percent. In parallel, Levy (2018) finds that firms in the special regime have little incentive to grow, even if they have high TFP_i. In fact, for those that are close to but below the threshold, even a 30 percent increase in sales pushes them into the general regime and thus results in lower after-tax profits.

3.5.3. The size of firms

Tables 1 and 2 document that self-employment and employment in small firms is very large in Latin America, and that the size distribution of firms is extremely skewed towards smallness. Many policies in $\mathbf{E}\{.\}$ contribute to these outcomes, some unrelated to social protection, including problems with contract enforcement as discussed below. However, four listed in Table 10 directly stem from social protection policies.

[45] To avoid this, some firms divide into separate legal entities that are below the thresholds, or, even if they are above, rather than expand when demand increases, subcontract to firms below the thresholds to lower their costs. Buying from a low-productivity firm may be cheaper than producing in a high-productivity one.

Table 10: Social Protection and Firm Size

Policy	Mechanism
Enforcement of (L_{CSI} + L_{JS} + L_{MW}) laws	Probabilities of being fined increase with firm size
Exemptions and ambiguities in (L_{CSI} + L_{JS} + L_{MW}) laws in firms without relations of dependency and subordination	Organizational obstacles for firm growth. Hiring workers under relations of subordination and dependency increases flow and contingent labor costs
Special tax and social insurance regimes	Smaller firms face lower labor costs and tax rates, and may not be allowed to sell to other firms
Minimum wages	Increasing with firm size

Source: authors' elaboration.

The first reason can be seen in the second row of Table 7. If, as it happens in Latin America, enforcement of (L_{CSI} + L_{JS} + L_{MW}) laws is concentrated in larger firms, then L_{ENF} increases with firm size. Because larger firms are more easily detected than smaller ones, informal firms stay small to be able to evade.

Although enforcement proportional to size is often mentioned as the main reason for the small size of informal firms, this is not the case in many countries in the region. As the third row of Table 7 reminds us, exclusions and ambiguities in the (L_{CSI} + L_{JS} + L_{MW}) regime allow some firms to be informal and perfectly legal. But these exclusions and ambiguities are not size neutral.

The one-person firms of the self-employed are the smallest possible firms, and in many countries they are exempt from (L_{CSI} + L_{JS} + L_{MW}) laws; or, when they are required to contribute, as in Argentina and Brazil, they can do so under very generous special regimes. But there are also countries, like Mexico, that exempt firms without dependent workers from (L_{CSI} + L_{JS} + L_{MW}) laws or—for practical purposes the same—where these laws are ambiguous and are replaced by tacitly accepted social norms, like in Colombia, the Dominican Republic, Honduras, Ecuador and Peru. These firms, however, tend to be small because it is difficult to coordinate many workers when there is no boss; problems of shirking or coordination grow with the number of workers (Levy, 2018). In fact, some could be larger if they established a hierarchical structure, but this would imply relations of dependency and subordination with their workers. However, these relations would put them in the second row of Table 7, at the risk of having to deal with the authorities enforcing (L_{CSI} + L_{JS} + L_{MW}) laws; the costs and contingent liabilities associated with that are too high for many firms to bear, and thus they stay small.

The third reason is equally important and is associated with special regimes, as discussed above. The fourth, finally, has to do with minimum wages that increase with the size of firms, as exemplified by Honduras and the Dominican Republic. Here the effect is very similar to that of special regimes: unit labor costs jump if the firm grows.

The size distribution of firms in Latin America reflects the formal-informal composition of the economy. The larger the informal sector, the lower the average size of a firm, and the more the size distribution of firms will be skewed in the direction of smallness.[46]

[46] These issues are usually studied under the rubric of "size-dependent policies," which in general have been shown to be quite costly in terms of **TFP** (Guner et al. 2008). In Latin America, however, these issues are more complex because size-dependent policies interact with the contractual structure of firms and a context of imperfect enforcement.

3.5.4. Digression: imperfect contract enforcement and firm size

The institutions charged with enforcing commercial and credit contracts are critical elements of countries' $\mathbf{E}\{.\}$s, captured in the "realm of market conditions" M(.) in relation (5). *Mutatis mutandis* across the region, these institutions work imperfectly, and this is another factor, in this case outside the domain of social protection policies, that affects firm size.

If banks face uncertainty or high legal costs to impound collateral when firms default on their loans, they will issue less credit. For many reasons, smaller firms are more likely to be in this situation, and thus less likely to get credit. In parallel, if firms face high legal costs enforcing commercial contracts, they will avoid arm's length transactions with unknown clients and engage only with those that are close by, known and trusted. Dispute settlement through the courts can then be replaced by dispute settlement through social norms. This may limit firm growth, but will reduce risks and increase expected profits, all the more so if being small allows the firm to avoid the (L_{CSI} + L_{JS} + L_{MW}) regime and perhaps benefit from special regimes.

Many small informal firms in Latin America are not incorporated as a legal entity, where the assets of the firm are distinct from the assets of the owner. This has drawbacks when assets have to be pledged as collateral for loans, but also advantages. Registering as firms increases the likelihood that those enforcing (L_{CSI} + L_{JS} + L_{MW}) laws will find them, something that could be catastrophic. Better to stay a small, unregistered enterprise, all the more so if the procedures to register are cumbersome, involving costly notaries and high transaction costs (De Soto, 1989). Yes, incorporating as a firm could have the advantage of using the courts to enforce contracts with clients and suppliers, but if the institutions charged with those tasks are slow, corrupt or unpredictable, why bother? Again, better to stay on the sidelines of the law, and transact only with trusted partners, settling disputes through social norms. The legal world is for large firms that can afford lawyers, accountants and notaries, and have friends in the right places when legal disputes occur.

Put differently, weak contract enforcement is yet another factor that keeps firms small, in addition to those discussed in Tables 7, 9 and 10. Better contract enforcement would improve access to credit and allow firms to diversify their client base and grow. Because larger firms would be more likely to have dependent workers and could not evade so easily, they would offer more formal jobs and improve social protection outcomes. In Latin America, weak contract enforcement conspires with flawed social protection policies to allow small, informal and unproductive firms to survive, hurting workers' access to social protection and hurting **TFP**.[47]

[47] Erosa (2001) shows that intermediation costs vary greatly across countries and have a large impact on firm size and self-employment. Quintin (2008a) shows that imperfect contract enforcement reduces firm size and increases the informal sector, and that quantitatively this effect is as relevant as imperfect tax enforcement. Quintin (2008b) shows that imperfect enforcement can account for the large differences in the size of manufacturing establishments between the United States on one hand, and Argentina and Mexico on the other. Laeven and Woodruff (2007) and Dougherty (2014) show that in Mexico, all else equal, states with weaker contract enforcement have smaller and more unproductive firms.

3.5.5. Interactions between size and formal-informal composition

Firms in Latin America perform in the context of Tables 7, 9 and 10. For large ones, Table 7 is a minor issue and Tables 9 and 10 are irrelevant. They are in the first row of Table 7 and considerations of labor costs may be far from the top of the agenda; they are concerned with innovation, international competition and so on. For medium size ones, Tables 9 and 10 may also be irrelevant, although Table 7 may matter more as, at the margin, these firms may be cheating or juggling contractual arrangements to lower labor costs. But for small firms, Tables 7, 9 and 10 are central; those with low TFP_i may only exist because of them, and those with higher TFP_i may not grow as a result of them.

Although there are commonalities, outcomes from Tables 7, 9 and 10 vary across countries, reflecting the interaction of many policies and parameter values: magnitudes of L_{CSI}, L_{JS} and L_{NCSI} and valuations v_{CSI}, v_{JS} and v_{NCSi}; values of w^{min} (relative to the wage distribution); effectiveness of enforcement L_{ENF}; differences in tax treatments from special regimes; and exemptions, ambiguities and social norms regarding (L_{CSI} + L_{JS} + L_{MW}) laws.

The average value of L_{CSI} in Latin America is 27 percent of the wage, but in Honduras it is 16 percent and in Argentina, 42 percent. L_{CSI} includes health almost everywhere, but not in Brazil. In Peru, L_{CSI} increases with the size of the firm, as shown in Table 8, while in Mexico, L_{CSI} is the same for all firms, but higher for lower-wage workers (35 percent for those earning one minimum wage and 21 percent for those earning 10). As noted, L_{CSI} includes cross-subsidies to informal workers for programs like early child development in Colombia, education and the national institute for agrarian reform in Brazil, and labor training in Peru.

w^{min} is set at the 19th percentile of the wage distribution in Argentina, the 35th in Brazil, the 16th in Chile, the 50th in Colombia and Peru, and the 25th in Mexico. In Honduras, w^{min} increases with the size of the firm but the average is set at the 60th percentile. Peru's job stability regulations make it almost impossible for firms to dismiss workers with permanent contracts, but Chile's allow it if there is an economic need to do so. In Ecuador, L_{NCSI} programs only provide health services, but in Mexico they also provide childcare. In Ecuador and Peru, poor workers lose access to their CCT if they get a formal job. There is also substantial variation in pension regimes. Brazil's is PAYG and Chile's is defined contribution, while some countries like Colombia, Peru and Honduras have both; and there are also large differences in rules to qualify for benefits. In parallel, there is also large variation in special tax regimes; in Argentina, these are mostly focused on the self-employed, but in Brazil, Ecuador and Peru, they extend to firms of various sizes.

Put differently, Tables 7, 9 and 10 are a vehicle to capture the effects of social protection policies on the formal-informal composition of the economy and the size of firms, but there is heterogeneity in the region in terms of regulations and parameters, which in turn accounts for heterogeneity in the relative importance of the various policies captured in these tables.

In this context, there are two observations on the empirical evidence referred to in the discussion of these tables. The first is that by and large it is obtained from studies that consider only one policy at a time, i.e., studies that analyze only the effects of a special tax regime, a change in the minimum wage, the expansion or creation of a specific program in L_{NCSI}, a reform in only in one component of L_{CSI}, or a specific universal or targeted transfer in T_{TRN}. This is quite understandable, even desirable, since this is the best way to precisely

identify and measure their impacts. That said, what matters for **TFP** is the joint effects of all the policies in Tables 7, 9 and 10. To identify their joint effects, the ideal counterfactual would be to ask: what would be the composition of employment and the number, size and contractual structure of firms without any of these policies? Answering this question is very much a work-in-progress, although our presumption is that once the interactions between all policies in Tables 7, 9 and 10 are considered, their quantitative impact would be quite large and provide a different perspective from that obtained from studies that only study one policy at a time.[48]

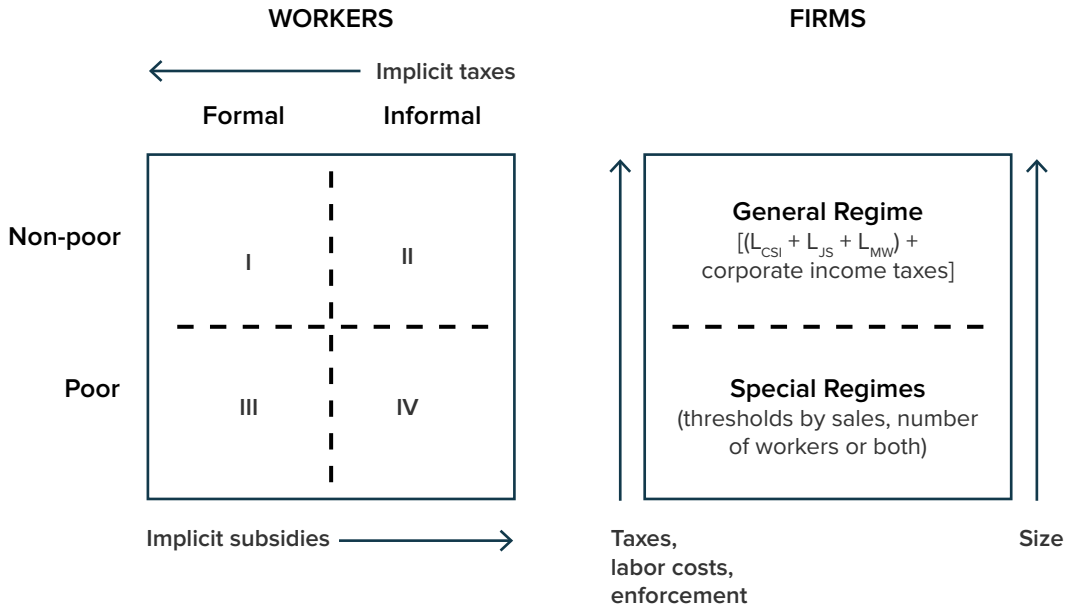
The second observation is that care needs to be exercised in extrapolating the results from one country to another, even if only a single policy is considered. For instance, a study for Peru that finds that lowering L_{CSI} contributions has no impact on formal employment does not imply that the same would hold in Mexico. The fact that in the first case the minimum wage cuts off at the 50th percentile of the wage distribution may nullify the impact lower L_{CSI}. But that same policy could have a different impact in Mexico, where it cuts at the 25th percentile. In any event, the impacts depend on the value of v_{CSI} in each country. Similarly, the fact that a study finds that adding or expanding a non-contributory program in L_{NCSI} has no impact on informal employment in Honduras does not imply that the same would hold in Ecuador since in Honduras other elements in Tables 7, 9 and 10 may have already compressed formal employment, implying that the effect of an additional policy is marginal.

Undoubtedly, the results from studies of any policy in a given country are very valuable and enhance our understanding of how that same policy could work in a different country. But policies have to be considered in the context of each country's L(.) and T(.) or, more correctly, in the context of each country's $\mathbf{E}\{.\}$. The empirical evidence referred to above should be viewed in this light. It hopefully illustrates the impacts of some of the policies captured in Tables 7, 9 and 10 on the behavior of firms and workers. But it should not be interpreted as an indication that each policy will have the same effect in every country in the region.

Graph 12 summarizes the discussion so far. The left panel refers to workers where, as in Graph 2, we classify them by labor status and income levels. The right one refers to firms, with those subject to the general tax and (L_{CSI} + L_{JS} + L_{MW}) regimes in the upper panel and those under a myriad of special regimes in the lower one.

[48] Antón et al. (2012) simulate the joint impact of L_{CSI} and L_{NCSI} programs on the composition of employment in Mexico, and find that in their absence, self-employment would fall by 31 percent, with a concomitant increase in salaried employment, suggesting that jointly they have a very large impact. That said, their results are far from satisfactory because their model does not capture impacts on firm size and ignores L_{JS} and special tax regimes, which are quite relevant in Mexico.

Graph 12: Workers, Firms and Social Protection Policies



Source: authors' elaboration.

Together with Tables 7, 9 and 10, Graph 12 hopefully illustrates the more abstract point made in chapter two in the context of relation (5): social protection policies impact resource allocation. In a nutshell: they segment workers and firms by a multitude of criteria and, at the end of the day, tilt resource allocation towards informality and smallness.

3.5.6. Social protection policies, resource allocation and productivity

Recall from relation (4) that **TFP** is the average of the productivities of individual firms, TFP_i, weighted by the share of resources allocated to each, a_i. It is illustrative to re-write it as:

$$(7) \quad \mathbf{TFP} = [\, a_1 TFP_1 + ... + a_{nf} TFP_{nf} \,] + [\, a_{nf+1} TFP_{nf+1} + ... + a_{(nf+ni)} TPF_{(nf+ni)} \,]$$

where nf and ni is the number of formal and informal firms respectively, and n = nf + ni the total number of firms. Clearly, ni > nf by a very large margin.

We reiterate that from the point of view of **TFP**, the existence of small informal firms need not be problem *per se*. As noted, if the TFP_is of all firms were the same, ni, nf and the a_is in relation (7) would be irrelevant. However, the problem is that all the available empirical evidence indicates that the TFP_is of informal firms are substantially lower than those of formal ones. More precisely, the problem is the combination of many resources channeled to informal firms together with the large productivity differences with formal firms.

Each individual informal firm absorbs an almost negligible amount of capital and labor (i.e., each has a miniscule a_i), a situation that at times leads one to think that these firms are irrelevant from a productivity point of view. But this is a mistake: what matters for **TFP** is not the a_i of any individual informal firm, but their sum. From this perspective, it is clear that these

firms matter substantially for **TFP** in Latin America, as shown by the figures on the distribution of employment and size of firms in Tables 1 and 2, and the formal-informal composition of employment in Table 3.

Informal firms are less productive than formal ones for many reasons. Reduced interest in exploiting economies of scale implies that the technologies deployed from set $\mathbf{T}$ in relation (5) are less sophisticated than those deployed by formal firms, even if they produce the same or similar goods. They can use technologies where the marginal revenue product of labor is below $[w^{min} + (L_{CSI} + L_{JS})]$ because they are exempt from these laws or because they evade them. Their sometimes illegal status obstructs access to credit from commercial banks, and they rely instead on short-term credit suppliers. They are often managed by individuals with little entrepreneurial talent—individuals who, if the parameters of Tables 7, 9 and 10 were different, would likely be workers in larger firms rather than manager-owners of tiny firms (so that some a_is in relation (7) would be zero, making the corresponding TFP_is irrelevant). When they are managed by talented individuals and have relatively higher TFP_is, the parameters of these tables conspire to keep them small since growth reduces after-tax profits.

Graph 3 highlighted the interdependence between social protection and firms. In Latin America, social protection outcomes are far from what is desired because there is too much self-employment and too many small and unproductive informal firms. And there is too much self-employment and too many small and unproductive informal firms because social protection policies bias the allocation of resources towards the informal sector.

3.5.7. Formalization programs and productivity

At times, countries in Latin America establish formalization programs based on special regimes to facilitate firm registration with the tax authorities and worker access to contributory programs. Here we use relation (7) to consider their implications for **TFP**.

Assume a previously informal firm formalizes, taking advantage of a special regime, so that now the formal-informal composition of firms changes, that is, n = (nf + 1) + (ni – 1). Does increased formality imply higher **TFP**? Relation (7) makes clear that unless the a_i and TFP_i of the formalized firm changes, **TFP** will not change; the economy will have formalized with no implications for productivity.

To pursue this point, it is useful to group firms into three sets. First, there are those that pay income taxes at normal rates and all the flow and contingent costs associated with the $(L_{CSI} + L_{JS} + L_{MW})$ regime; call them standard formal firms. Second, there are firms in special regimes, where workers get some protections from $(L_{CSI} + L_{JS} + L_{MW})$ programs but, critically, this group faces very different tax and social insurance burdens than standard formal firms; call them formalized informal firms. And third, there are informal firms that do not participate in formalization programs; call them standard informal firms. Clearly, what matters for **TFP** are firms' a_is and TFP_is, not their labels. If the special regime that supports the formalization program is very generous, labor costs and tax burdens of formalized informal firms will be almost the same as when they were standard informal firms; their a_i and TFP_i will not change. As a result of the formalization program, the economy will now have one less informal and unproductive firm, and one more formal but equally unproductive firm; **TFP** will not change.

tIn other words, from the perspective of **TFP**, formalization programs can be quite misleading because they provide firms with a veneer of legality, while changing almost nothing except

their label. In fact, these programs could actually hurt **TFP** if formalized informal firms expanded, taking advantage of their newly acquired status.

Consider two examples. First, in Brazil in 2012, formal employment was 63.7 percent of the labor force, with 54.2 percent under the general regime and 9.5 percent under two of its special regimes (the Simplified System of Contributions and the MEI). In 2019, formal employment was almost the same, 63.8 percent of the labor force, but now 50.8 percent was under the general regime and 13 percent under its special regimes.[49] In other words, even though labor formality did not change, its composition did, in a direction most likely unfavorable to **TFP**. Second, Levy (2018) computes the profits of firms that qualify for Mexico's special regime and finds that they are almost the same as without paying anything, i.e., the special regime allows firms to formalize with practically no impact on their after-tax profitability. In contrast, profits would fall substantially if they were taxed under the general regime, implying that many would exit the market or be significantly smaller.

It is useful to return to the discussion of the words formal and informal in chapter two. These words matter for social protection because they identify which workers are protected against what risks and are covered by regulations on minimum wages and job stability. These words are not relevant for **TFP**. From the point of view of productivity, what matters is that low-productivity firms exit the market, that high-productivity ones grow and that those entering the market have higher productivity than incumbents.

The association between low productivity and informality is understandable because, as discussed, low-productivity firms are usually informal, i.e., firms with low TFP_is are those outside the domain of the ($L_{CSI} + L_{JS} + L_{MW}$) regime. But it is important to emphasize that these firms exist because they face different labor costs and tax burdens than formal ones. Unless formalization programs change these determinants of firm behavior, they will not increase **TFP**, and could in fact hurt it if they further distort the process of firm entry, exit and growth.

3.5.8. The misuse of human capital

We now discuss briefly the second channel by which social protection policies can hurt **TFP**: the misuse of human capital. Firms using less sophisticated technologies require fewer engineers and computer programmers. If they are on the border between legality and illegality, they also need fewer lawyers and accountants. And if they innovate little, they also need fewer mathematicians, biologists or chemists. The point is this: large informal sectors depress the demand for workers with more schooling. Latin America's investments in education and training programs fail to bring the expected returns because there are too few firms that require workers with sophisticated skills, and too many that do not. This can depress the incentives to invest in human capital.[50]

Many informal firms die two or three years after they start. However, when this happens, they are replaced by similar ones. These firms have higher exit rates than formal ones, but also

[49] The comparison begins in 2012 even though the Simplified System started in 2007 and the MEI in 2009 because there was a break in the employment series in 2011.
[50] Bobba, Flabbi and Levy (2017) model workers' schooling decisions based on whether they expect to have a formal or informal job when they enter the labor market and show that informality depresses investments in schooling.

higher entry rates. Large firm churning is reflected in large worker churning, a situation that both lowers firms' incentives to invest in worker training and lowers workers' opportunities to learn on the job and acquire abilities. This is another factor that lowers **TFP**. Large, useless churning explains why the returns to workers' experience is lower in Latin America that in other regions.[51]

Many other policies in $\mathbf{E}\{.\}$—unrelated to social protection—also matter for these outcomes. But the point here is that the matches between firms and workers of different schooling and abilities, and firm and worker churning, are also influenced by the laws and policies in L(.) and T(.) that jointly conform social protection in Latin America. As with Tables 7, 9 and 10, the magnitudes of the effects and the relative weight of the specific channels vary between countries but, *mutatis mutandis*, the result is that these policies have three consequences for **TFP**:

- they may impede workers from fully benefiting from their investments in schooling prior to their entry into the labor force;
- they may diminish workers' opportunities to enhance their human capital while working; and,
- they may depress workers' incentives to invest in human capital.

3.6 Trade-offs Between Social Protection and Productivity Objectives

This section discusses the dilemmas between social protection and productivity objectives associated with the region's social protection policies.

3.6.1. Coverage of insurance

Table 7 helps to identify trade-offs between coverage of insurance and **TFP**. As discussed, many workers are excluded from the coverage of contributory programs. If there were no non-contributory programs, such that $L_{NCSI} = 0$, informal workers would be left without any insurance, something that is clearly undesirable from a social point of view. But it is also clear from rows two and three in that table that, because there are no subsidies to informality, the share of resources channeled to the informal sector would be smaller and **TFP** higher.

It is sometimes said that universal coverage of social insurance coverage could be reached in Latin America by extending L_{NCSI} programs to all informal workers, while formal ones are covered by L_{CSI} programs. Further, inequities could be eliminated if L_{NCSI} protected against the same risks as L_{CSI} and offered services of the same quality. But a moment's reflection shows that the implications of universalism through "formal in L_{CSI} and informal in L_{NCSI} with $L_{CSI} = L_{NCSI}$" are very different from those of universalism through "all in L_{CSI}". Aside from the fact that risk pooling and aggregate savings would be different, what would happen to **TFP** if $L_{NCSI} = L_{CSI}$?

[51] Lagakos et al. (2018) find that the returns to experience are substantially higher in developed countries than in Brazil, Chile and Mexico (the three countries from the region included in their sample). Bobba, Flabbi, Levy and Tejeda (2020) find that workers accumulate less human capital while informally employed. Levy (2018) shows that the returns to experience in Mexico are flat.

Job stability regulations also generate trade-offs. Covered workers enjoy permanent access to contributory programs, but since firms deploy various strategies to avoid them, many others are left out. The implied **TFP** losses are hard to measure because practices vary. In all countries, dismissals require severance payments. In a few, changes in demand and technical change are recognized as just causes of dismissal, but in most they are not. Further, severance pay varies across countries depending on seniority and other factors.

Da-Rocha et al. (2016) measure the loss of **TFP** caused by job stability regulations. They estimate that if severance pay equals one year of salaries, **TFP** falls by 4.2 percent and if it equals five, by 20.6 percent. These amounts are in the range observed in the region. However, their results are obtained under the assumption that firms can dismiss workers freely as long as they cover the associated costs. But this is not the situation in Latin America, with the possible exception of Chile. The fact that job stability regulations are more complex suggest that the trade-off between them and **TFP** is steeper than what these numbers suggest.

3.6.2. Poverty

Some combinations of parameters in Table 7 can constitute almost insurmountable barriers for poor workers to enter formality. First, the implicit tax on formality may be higher because they face higher contribution rates for L_{CSI} programs (like in Mexico). $[w^{min} + (L_{CSI} + L_{JS})]$ may be above the value of their marginal revenue product given the TFP_i of the firms that could hire them, and even if they are self-employed and want to contribute to L_{CSI} programs, costs are too high when their earnings are below w^{min}. Their v_{CSI} may be lower because they transit more often between formality and informality than higher-income workers, and thus they are less likely to get a contributory pension. They may also live in rural or marginal urban areas where L_{CSI} health facilities are scarcer.

The subsidy to informality is also proportionately higher for poor workers because L_{NCSI} programs of fixed monetary value (like flat pensions or child allowances) represent a relatively larger share of their earnings than for non-poor workers, and because in some countries these programs are focused only on poor workers.

Whatever the combination, excluding poor workers from $(L_{CSI} + L_{JS} + L_{MW})$ programs makes it harder to reduce poverty. Raising w^{min} is ineffective because the majority of poor workers are self-employed or work in informal firms. Increasing cash transfers through targeted programs works, but at a cost: if poor workers get a better job where their income exceeds $\mathbf{Y}^{cutoff}$, they lose the transfer and may be worse off as a result. For them to be better off, the additional income needs to exceed the value of the transfer, and even then, the implied marginal tax rate may be quite high. As a result, increasing targeted transfers can reduce $P(\mathbf{Y}, \mathbf{Y}^{cutoff})$, at the cost of trapping poor workers in low-productivity jobs and diminishing their contribution to **TFP**.

Some CCTs seek to avoid this trap through a lifecycle approach. If transfers are based on the number of children in the family and conditioned on their school attendance, they gradually fade away as youth finish high school. The expectation is that their enhanced human capital will help them get a more productive formal job, making future transfers unnecessary. But if entry into such a job is blocked by the parameters of Tables 7, 9 and 10, transfers need to be permanent, their improved human capital notwithstanding.

3.6.3. Inequality

Recall that taxes and transfers reduce inequality if G(**Y** – **t**) < G(**Y**). Importantly, however, **t** not only determines the difference between market and disposable incomes; it also impacts firms' and workers' behavior and thus market incomes.

This point is central, and it is useful to make it with more precision. Assume that taxes and transfers are $\mathbf{t}^0 = t_1^0, t_2^0, ..., t_n^0$. In policy discussions in Latin America it is sometimes asserted that the Gini coefficient falls from G($\mathbf{Y}^0$) to G($\mathbf{Y}^0$- $\mathbf{t}^0$) as a result of taxes and transfers $\mathbf{t}^0$. But this assertion needs to be interpreted with care because without taxes and transfers $\mathbf{t}^0$, the Gini coefficient would not be G($\mathbf{Y}^0$); it would be another number.

Put differently, taxes and transfers $\mathbf{t}^0$ are part of the environment $\mathbf{E}^0\{.\}$ that through relation (5) result in market incomes $\mathbf{Y}^0$ and Gini coefficients G($\mathbf{Y}^0$) and G($\mathbf{Y}^0$ – $\mathbf{t}^0$). However, if taxes and transfers were different, say $\mathbf{t}^1$ and not $\mathbf{t}^0$, market incomes would no longer be $\mathbf{Y}^0$. They would be $\mathbf{Y}^1$ because the environment is now $\mathbf{E}^1\{.\}$. The Gini coefficient would not change from G($\mathbf{Y}^0$ – $\mathbf{t}^0$) to G($\mathbf{Y}^0$ – $\mathbf{t}^1$), but from G($\mathbf{Y}^0$ – $\mathbf{t}^0$) to G($\mathbf{Y}^1$ – $\mathbf{t}^1$).

Now, to reduce inequality, one can expand non-contributory and targeted programs, changing $\mathbf{t}^0$ to $\mathbf{t}^1$ so that:

$$(8)\ G(\mathbf{Y}^0 - \mathbf{t}^0) > G(\mathbf{Y}^1 - \mathbf{t}^1)$$

But, as discussed, as non-contributory and some targeted transfer programs expand, resources shift towards the informal sector. G(**Y** – **t**) falls, but so does **TFP**.

In Latin America, the taxes and transfers in Table 7 usually account for around 6 percent of countries' GDP, sometimes more.[52] This is a large number that inevitably affects the behavior of all. Absent taxes and transfers worth 6 percent of GDP, market incomes would be very different. In other words, taxes and subsidies **t** not only change the distribution from market to disposable income, they also affect **TFP**—a point that relation (5) underlines, but that is not always kept in mind in discussions of income inequality focused on relation (8).

Studies find that G(**Y - t**) fell in Latin American countries between approximately 2002 and 2018. These studies decompose this fall into two components, one derived from changes in **t** and another one from a compression of the earnings distribution, i.e., a narrowing of the differences in market income between high- and low-income workers. Although numbers vary across countries, the general finding is that the latter component made a larger contribution to reducing inequality than the former; see López-Calva and Lustig (2010), Messina and Silva (2019) and IDB (2020a).

These results highlight the need to understand the reasons behind the compression of the earnings distribution. One reason could be increases in the real value of w^{min}, which may or may not negatively impact **TFP** depending on changes in unemployment and the formal-informal composition of employment. But earnings compression can occur for other reasons, as discussed by Messina and Silva (2019).

---

[52] Contributions to social security programs are usually around 3 percent of GDP and spending in L_{NCSI} programs between 1 and 2 percent (Izquierdo et al., 2018). Transfers programs in T_{TRN}, excluding education, would add another 1 to 2 points.

Acosta et al. (2019) document the evolution of wage differentials and the supply of workers by educational level for 16 Latin American countries over the period 1991-2013. They find that the returns to secondary education fell over time and, in contrast, that the returns to tertiary education display a remarkable changing pattern common to almost all economies, with significant increase in the 1990s, strong fall in the 2000s and a deceleration of that fall in the early 2010s. Since the supply of both education groups increased over the period (relative to low-skilled workers), the authors conclude that supply-side factors seem to have limited explanatory power relative to demand-side factors in accounting for changes in the wage gap between workers with tertiary education and the rest.

Levy and López-Calva (2020) focus on demand-side factors in Mexico during 1996-2015, a period characterized by macroeconomic stability and a constant real minimum wage. They document increasing misallocation of resources and a large and growing presence of informal firms. They also document that these firms require fewer workers with more schooling relative to formal ones (presumably because they use simpler technologies). In this context, they show that growing firm informality depressed the relative demand for workers with more schooling; in turn, this compressed the wage distribution, lowering G(**Y**). But in this case one could say that inequality fell for the wrong reasons. From the point of view of **TFP**, it would be better if, because there are fewer informal firms and more formal ones, wage differentials widened, even if this implied a higher G(**Y**).

Returns to education are an important determinant of G(**Y**). Ignoring the minimum wage and short-run macroeconomic fluctuations, their evolution is often explained as the outcome of a race between technology and schooling. If technological change displaces low-skilled workers and requires more skilled ones, returns will increase, unless this is offset by an increased supply of skilled workers (Katz and Murphy, 1992). The outcome of the race determines the direction of change in these returns and their impact on G(**Y**). The point to be made here is that the technologies deployed in Latin America are a mix of those used by formal and informal firms. This means that the demand for workers of different skill levels depends partly on the formal-informal composition of the economy. In other words, new technologies can evolve in the direction of requiring more skilled workers, but this trend is only reflected in the labor market if firms actually use them. If a significant subset of firms continues to use older technologies, the nature of the race between technology and schooling is altered. This point is irrelevant in developed countries but may be very relevant to some countries in Latin America, as exemplified by the case of Mexico.

Because Latin America is such an unequal region, reducing income inequality looms large in the policy agenda. However, the policies implemented to do so matter substantially. The segmentation of social insurance makes it more difficult to lower G(**Y** – **t**) without impacting **TFP** because in that context some policies—like raising contributions for contributory programs or expanding non-contributory ones—may negatively affect margins of behavior that are central to **TFP**. Of course, this is not true of all policies. Income inequality can be reduced through more effective income taxation, through more vigorous competition policy (the latter in fact helps to increase **TFP**) or, as discussed in the next chapter, the restructuring of social protection. The trade-offs are completely different, as is the political economy.

All countries face trade-offs between social protection and productivity outcomes. What is distinct in Latin America is that they are steeper because they occur in economies with large informal sectors. In this context, it is useful to remember that informality does not derive from the fact that countries in the region speak Spanish or Portuguese nor, for that matter, from the fact that the continent is surrounded by the Atlantic and Pacific oceans. Countries are

not immutably divided into formal and informal sectors by the languages that they speak or by their location on the planet; they are divided by their $\mathbf{E}\{.\}$s, with social protection policies playing a prominent but not exclusive role. If these policies were different, the trade-offs discussed here would be too, and the region would be better poised to pursue its social protection objectives without undermining the foundations of its prosperity.

3.7. Social Protection and the Strength of Rule of Law

The strength of the rule of law—the degree of compliance with a country's laws and regulations given citizens' predispositions to observe them and the effectiveness of institutions enforcing them—matters for social protection outcomes. Strong rule of law means that firms and workers comply with their obligations with regards to social insurance, that minimum wages are observed, and that workers get the compensations they are entitled to when they are dismissed. The causation from rule of law to social protection outcomes is clear and needs no further discussion.

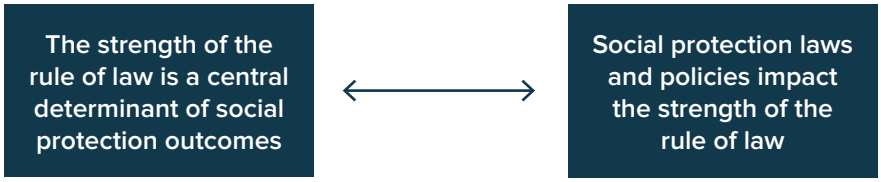
In this section, however, we argue that social protection laws and policies also impact the strength of the rule of law. This two-way interdependence is highlighted in Graph 13.

The (L_{CSI} + L_{JS} + L_{MW}) regime is the keystone of the region's social protection architecture—a regime begun in the 1930s in South America and extended to Central America, the Dominican Republic and Mexico in the 1940s and 1950s. It is impossible to understand social protection ignoring the context of when these laws were born. They were often the means to end political conflicts and, at times, military ones. They were partly a mandate for immediate action, but partly a *desideratum*—an image of the societies that Latin America aspired to become.

(L_{CSI} + L_{JS} + L_{MW}) laws were born with the relation of dependency and subordination at center stage. Social insurance was conceived as a right of workers, not citizens—and for that matter, not all workers. This is the original sin of social protection in Latin America. These laws could not produce universal coverage and, from this perspective, the seeds of Latin America's segmented social insurance system, and the emergence of non-contributory programs decades later, were there from the beginning.

Since then, laws on L_{CSI} have evolved—a health reform here, a pension reform there. In parallel, some countries like Argentina, Brazil, Colombia and Peru have extended L_{CSI} to include societal needs unconnected to insurance: child allowances, early child development, education, housing, labor training. In others, like Mexico, they have been used as a vehicle to raise revenue for subnational governments.

Graph 13: Interdependence between Social Protection and the Rule of Law



- Are firms sanctioned if they hire workers under relations of dependency and subordination but do not comply with (L_{CSI} + L_{JS} + L_{MW}) laws?
- Are self-employed workers sanctioned if they are required to contribute to L_{CSI} programs but fail to do so?
- Is enforcement proportional to the size of the firm or income level?
- Do enforcement agencies have the incentives to enforce laws fairly and evenly?
- How effectively do labor tribunals solve firm-worker disputes? Are rulings complied with?
- Do laws specify clearly the criteria under which firms and workers should participate in L_{CSI} programs?
- Can these criteria be easily verified?
- Do L_{NCSI} programs provide free or subsidized benefits to workers that should be participating in L_{CSI} programs?
- Do laws have thresholds or discontinuities that generate sharply distinct outcomes?
- Do laws treat individuals that perform similar tasks differently? Do they favor arbitrage across different tax or contributory regimes?

Source: authors' elaboration.

Laws on L_{JS} have evolved as well, complemented with jurisprudence from labor tribunals and constitutional courts. Some countries, like Peru, have made firing permanent workers almost prohibitively expensive. Firms could withstand this under the protective mantle of import substitution industrialization but can hardly withstand it as economies have integrated into the world economy. Still other countries, like Brazil, Colombia, Honduras and, more recently, Mexico, have sought to strengthen their redistributive efforts by changing L_{MW} laws to increase minimum wages.

Reforms have been the response to a myriad of country-specific circumstances. But, by and large, the original (L_{CSI} + L_{JS} + L_{MW}) regime still stands: social insurance as a right of workers and, in most countries, of only a subset of them.

After the lost decade of the 1980s, the shortcomings of this regime were evident: more than half of the region's workers, including most of the poorer ones, were excluded from social insurance. But rather than coming to terms with the regime's underlying flaws, countries opted to bypass them. The overwhelming need to expand the coverage of social protection was attended through a parallel regime: a hybrid of programs in L_{NCSI} and T_{TRN} under an *ad hoc* combination of new laws, presidential decrees, or programs in countries' budgets, and a myriad of rules granting or denying access on the basis of this or that. Programs appear and disappear as presidents enter and exit, or change names or rules as political winds shift. The result is that workers' rights to social protection are erratic and capricious. Benefits like health and pensions are sometimes legal entitlements, and sometimes gifts from the government. In the second case, their legitimacy tends to be highly politized as benefit level changes are determined in partisan debates.

Firms are also affected by these parallel regimes. The persistence of (L_{CSI} + L_{JS} + L_{MW}) laws implies the persistence of burdens on the relation of dependency and subordination. Firms can escape these burdens by avoiding this relation; in fact, many have flourished because they do so, at the cost of changing their contractual structure. Others escape them through evasion: staying small, taking refuge in special tax regimes, or through subcontracting, simulation and clever lawyers. But firms have to observe a plethora of regulations because if they grow, they risk losing the benefits of special regimes, or face higher minimum wages, or step on the edges of the relation of dependency and subordination. Because the interpretation or enforcement of (L_{CSI} + L_{JS} + L_{MW}) changes, their legal obligations are also erratic and capricious.

The strength of the rule of law is not independent of the content of the law. This is a critical observation because social protection laws in Latin America are complex, ambiguous and at times contradictory. It is sometimes asserted that the region's problem is the imperfect application of the law. If enforcement were strengthened, social protection would improve. This is true, but unfortunately the problem is more complex than that.

The problem is a complex web of laws to the point that at times some laws subsidize the evasion of others. This can be seen in Table 7. Firms in the second row evade (L_{CSI} + L_{JS} + L_{MW}) laws because enforcement is imperfect. But social protection policies reward this behavior. Through programs in L_{NCSI} or T_{TRN}, workers get social benefits even if the firms hiring them evade taxes and contributions. Clearly, if these programs were not there, workers would be less willing to accept jobs without coverage of contributory social insurance, or at least they would ask for higher wages to do so. *De facto*, firms and workers are subsidized for breaking the law. In Argentina, all workers are required to participate in contributory health programs, including self-employed workers who can participate at highly subsidized rates through the *Monotributo* regimes, but the government spends at least 2 percent of GDP providing free health services to any citizen that requests them.

The problem is that laws are contradictory or ambiguous. In Honduras, the minimum wage law states that all contributions for L_{CSI} should be based on at least one minimum wage, but the Honduran Social Security Institute allows firms to register workers with wages below that. In Mexico, the law that sets non-contributory health benefits grants them to all workers who lack coverage from contributory programs, without distinguishing whether that lack of coverage results from firms' non-compliance with contributory programs or workers' exclusions from them because they are non-salaried.

The problem is that some social protection laws give large discretionary powers to government officials. In Colombia, the obligation of self-employed workers to contribute to L_{CSI} programs depends on whether their income is one peso higher or lower than the minimum wage. In Ecuador, firms with up to 10 workers may qualify for its special tax regime, but not if they hire 11. In Mexico, if sales are up to 1,999,999 pesos, firms pay a tax of 2 percent on sales, but if sales are 2,000,001 pesos, the tax is 35 percent on profits. Officials who decide whether workers earn one peso below or above the minimum wage, whether firm sales are two pesos above or below two million, or whether firms have 10 or 11 workers have substantial discretionary power because some of these thresholds are not easily observed or can vary over time. Even if corruption were not an issue, flawed assessments would be. Workers' earnings fluctuate, and so do firms' sales.

The problem is that under the (L_{CSI} + L_{JS} + L_{MW}) regime, the incentives to cheat are large. In Honduras, self-employed workers can avoid the obligation to contribute to L_{CSI} programs by

simply declaring that their income does not allow them to comply. A permanent contract in Peru gives workers the right to protection from L_{JS} laws; a sequence of temporary contracts does not. And in many countries, L_{JS} and L_{MW} laws are eluded by different contractual forms, a practice that is at times also used by government agencies that hire workers as independent contractors rather than as dependent workers. While firms and the government save money, workers' rights are trampled.

The problem is that special tax regimes for micro and small firms are also fertile grounds for deceit. Why not divide the firm into two legal entities, each paying taxes under the special regime, rather than a single firm paying taxes under the general regime? Transaction costs are higher, but so are after-tax profits. Moreover, why not do it, if minimum wages also increase with firm size? And if there are highly subsidized contribution regimes for self-employed workers, why not give workers a non-salaried contract, pay lower contributions—rather than hiring them as dependent workers—and share the difference in contribution rates? This would explain why, as in Argentina, 25 percent of the "self-employed" belong to firms with between 2 and 10 workers.

The problem is that many laws overlap with presidential decrees and administrative regulations from various ministries. When laws are complex, unclear and contradictory or subject to parameters that can vary constantly, the rights of workers and the obligations of firms are equally complex, unclear and contradictory—fertile ground for opportunism and corruption.

The result is what we all observe in Latin America everyday: millions of workers and firms in a limbo between legality and illegality; millions of firms evading, simulating or eluding; and millions of individuals on the edge between being a firm and being a worker, on the fuzzy border of the self-employed and the micro firm. When States fail to provide clear laws, or to enforce them efficiently, societies fill the vacuum with tacitly accepted norms. Put differently put, the large gap between the *de jure* and *de facto* status of Latin American firms and workers—with vast implications for the effectiveness of social protection—is not only a problem of weak rule of law; it is partly a problem of flawed laws.

There is another dimension to the relation between social protection and the strength of the rule of law. The fact that many informal firms and workers carry out their activities guided by tacitly accepted social norms, and not laws, depresses the demand for strong rule of law. Indeed, strong rule of law would impede them from continuing with their *modus vivendi*. However, if non-compliance is evident, ubiquitous and highly visible, this will likely undermine compliance in other areas. On the other hand, if strong rule of law means having to comply with an (L_{CSI} + L_{JS} + L_{MW}) regime that provides benefits that are well below its costs, who wants it? This is another mechanism by which social protection laws, with varying degrees of relevance across countries in the region, weaken the rule of law. For Latin America to enjoy stronger rule of law, it needs more individuals demanding stronger rule of law; this requires better social protection laws. Under the *status quo*, millions of firms and workers, while paying lip service to the law, will continue to run their affairs on the sidelines.[53]

[53] When Mexico was a colony of Spain, many laws were issued by the Spanish Crown for La Nueva España that were to be "observed but not complied with" as, famously, a viceroy is said to have said after realizing that they could not be applied.

3.8. Social Protection and Growth

This section discusses the relation between social protection policies and growth. To do this, we first briefly expand the framework developed in the previous chapter, in particular relation (5), to consider two factors that so far have been absent from the discussion but are central to growth in Latin America: conditions in the world economy and infrastructure.

3.8.1. Additional factors affecting growth

We do so in a very simple way. First, let $\mathbf{W}$ be the set of exogenously determined factors in the world economy that are relevant to Latin America, among them prices of agricultural products, metals and oil; growth in China and OECD countries; and international interest rates. These factors affect countries in the region differentially, but it is not necessary to provide any details here.

In turn, expand the set of institutions, laws and policies to include those associated with macroeconomic policy, which we label as the "realm of monetary and fiscal policy". These are represented by P(.) = (P_1, P_2,.., P_n), where P_i refers to the laws and policies governing the central bank; the exchange rate regime; fiscal rules; policies on the currency composition of public debt; and so on. Clearly, P(.) also varies by country, but again, the details are not relevant here. Our only objective is to capture a fourth component of the environment $\mathbf{E}\{.\}$, in addition to L(.), T(.) and M(.), that matters for resource allocation, **R**, aggregate productivity **TFP**, social protection, **SP** and incomes, **Y**.

Second, we also incorporate infrastructure in a very simple way, separating the capital stock K into two components: K_{firm}, allocated to individual firms that produce goods and services that are privately appropriated, as considered so far in the paper; and K_{inf}, allocated to infrastructure like transportation, telecommunications and water and sanitation.

With these definitions, and letting **GDP** be gross domestic product (= $\sum \mathbf{Y}_i$), we expand the basic relation driving this essay, relation (5), to:

(9) { [(I.H), (K_{firm}, K_{inf})]; $\mathbf{T}$; $\mathbf{W}$; $\mathbf{E}$ [L(.), T(.), M(.), P(.)] }

simultaneously determine [**R**, **TFP, SP**, **Y, GDP**]

Evidently, relation (9) is even more complex than (5), as outcomes are also affected by world conditions $\mathbf{W}$, by an additional set of institutions and policies captured by P(.), as well as by the availability of infrastructure, K_{inf}. Individuals making decisions with regards to occupational choices, firm size, contractual structure and so on now have to take into account, or at least are influenced by, factors stemming from the world economy, a broader set of domestic institutions and policies, and the availability and quality of infrastructure services derived from K_{inf}.

Policies in P(.) are the central subject matter of ministries of finance and central banks. They change often in response to the volatility of $\mathbf{W}$, or to domestically originated shocks. As is well known, in the past these policies have sometimes created instability, or exacerbated the volatility of $\mathbf{W}$, particularly when positive transitory shocks triggered more permanent increases in public spending (De la Torre & Ize, 2020).

However, after the debt crisis of the 1980s, the region strengthened key institutions in P(.), *mutatis mutandis* across countries, granting autonomy to central banks, adopting flexible exchange rate regimes, setting up stabilization funds, improving fiscal management and financial regulation, and so on. The upshot is that by the mid-1990s, most countries had achieved macroeconomic stability. This allowed them to focus on strategies to raise their long-term growth rate with the understanding that volatility in **W** would inevitably generate variations around a hopefully higher rate.

3.8.2. The region's growth strategy

The strategy consisted of two pillars:

- policy reforms to raise **TFP;** and,
- investments in human (I.H) and physical (K_{firm}, K_{inf}) capital.

Policy reforms in the first pillar centered on the "realm of market conditions", M(.). Many efforts were devoted to enhancing the region's participation in the world economy. Countries changed foreign trade and investment policies, M_{FTI}, through agreements among themselves (like Mercosur, the Central American Free Trade Agreement and, more recently, the Pacific Alliance) or with countries outside Latin America (like the North American Free Trade Agreement), as well as through unilateral changes to their own trade and foreign investment regimes. This was accompanied by efforts to take advantage of the growth of China and the dynamism of East Asia.

The expectation was that openness to foreign trade would weed out firms with low TFP_is and give those with higher TFP_is better access to export markets and imported intermediate inputs. Through trade, some domestic firms would connect with world value chains, and through backward linkages, indirectly connect other domestic firms as well. Firms could then exploit economies of scale and specialize in higher value products. Connectedness with world markets would shift the basket of goods produced towards those that required more complex technologies, offering greater possibilities of learning by doing. In parallel, foreign investment would bring in new firms using frontier technologies. In a nutshell, changes in M_{FTI} would trigger a dynamic process of resource reallocation leading to faster **TFP** growth.

However, the region's growth strategy recognized that trade and foreign investment policies were not enough. Spillover effects from one firm to another, and information, coordination or other types of market failures, meant that, left on their own, firms would not invest or innovate enough. Sometimes the TFP_i of one firm depended on the TFP_j of another because they both belonged in the same cluster and would be affected by each other's actions; and sometimes the TFP_i of both firms, or their investment plans, depended on critical inputs that only the government could provide, as an appropriate ecosystem was needed for firm development. Further, it was recognized that producing bananas was not the same as producing computers. Some activities have more potential for exploiting economies of scale, or for innovation, than others. These activities are usually in the manufacturing sector, but also in services and commerce. In some cases, information or coordination failures or failures in credit markets, may impede firms from engaging in them, putting Latin America at a disadvantage relative to other regions.

Thus, even if there were openness to international trade, the resource allocation **R** generated by the market would be suboptimal. In consequence, further efforts were required to raise

TFP under the label of industrial policies or productive development policies, gathered here under M_{PDP} in relation (9). These could take many forms: tax facilities to compensate firms for the risks of innovation; tailor-made regulations for specific sectors; transitory subsidies when spillover effects or positive externalities created gaps between social and private returns; credit from development banks to compensate for shallow capital markets; or the provision of public goods to facilitate coordination between firms (Hausmann, 2011; IDB, 2014).

Perhaps less forcefully, through competition policies (M_{COM}) countries also regulated mergers and combated monopolistic practices to ensure that gains in market share accrued to firms with high TFP_is and not to firms with friends in high political positions. Finally, there was widespread recognition of the importance of strong rule of law given its relevance to firm behavior, but it is probably fair to say that policies to improve the enforcement of commercial and credit contracts and, more generally, protect property rights (M_{ENF}), did not receive the attention that they deserved.

The second pillar of the growth strategy centered on enhancing the stock of human and physical capital. A more skilled labor force would allow firms to deploy sophisticated technologies that could not be used before because workers lacked special abilities. The capital goods purchased through new investments could be more complex because more knowledgeable workers could operate them. For these reasons, but also because they were considered essential for social mobility, investments in education took center stage.

In parallel, efforts to increase the stock of physical capital worked on two fronts. The first one was to stimulate investments in K_{firm}, which by and large were expected to be carried out by private firms, facilitated by a context of greater macroeconomic stability and complemented by policies in the "realm of taxes and transfers" T(.) like accelerated depreciation regimes and special treatments for firms when they located in some regions or engaged in activities considered to be strategic.

The second front was investments in infrastructure, an area where Latin America clearly underperformed *vis-à-vis* the rest of the world.[54] Because current expenditures were absorbing an increasing share of the budget—partly for social programs, particularly pensions—efforts centered on allowing private investments in areas that in some countries were previously reserved for the public sector, like energy and telecommunications, and on regulatory changes to foster public-private partnerships in projects in transport and water and sanitation (Benavides et al., 2020; IDB 2020b).

The mix of measures to promote growth varied across countries. Chile and Mexico relied heavily on changes to M_{FTI}, in the first case through a strategy of broad participation in world markets, and in the second with emphasis on trade and foreign investment agreements, particularly with its northern neighbors. Argentina and Brazil placed greater emphasis on measures in M_{PDP}. Countries like Honduras and the Dominican Republic intensively used tax subsidies and special export zones to foster tourism and promote maquiladora-type activities to increase K_{firm}. Many countries passed laws to facilitate public-private partnerships to increase K_{inf}. Efforts to expand the coverage and increase the quality of education were ubiquitous.

[54] Benavides, Guzmán and Perry (2020) compare annual investments in K_{inf} as a share of GDP in various regions for the period 2007-2017. The find that on average they were 7.7 percent in East Asia and the Pacific, 6.9 percent in the Middle East and North Africa and between 2.4 and 2.7 percent in Latin America.

Whatever the specific mix in each case, at a high level of abstraction the region's growth strategy could be represented by:

(10) { macro stability + [reforms in M(.) and T(.)] + more human and physical capital}

result in {higher **TFP** and faster **GDP** growth}

Of course, the strategy would be affected by shocks in $\mathbf{W}$ (positive and negative). But, as mentioned, these would be managed with appropriate macro policies in P(.). Thus, while these shocks were a source of concern (and absorbed a large share of policymakers' attention), at the end of the day the expectation was that the strategy summarized in (10) would raise the long-term growth rate, with inevitable but controlled fluctuations.

The most important fact about this strategy is that the first pillar failed. With some numerical differences across countries, by and large **TFP** stagnated. Growth resulted basically from the accumulation of human and physical capital, facilitated at times by favorable shocks in $\mathbf{W}$ (including growth in China). Table 11, taken from Fernández-Arias and Fernández-Arias (2021), decomposes the growth rate of per capita **GDP** for the period 1990-2017 into the sum of the growth rate of per capita factor accumulation, $[(I.H, (K_{firm}, K_{inf})]/N$, where N is total population, and the growth rate of **TFP**.

Table 11: Factor Accumulation, Productivity and GDP Growth
(Average growth rates, 1990-2017)

	Factor Accumulation (per capita)	TFP	GDP (per capita)
Argentina	1.26	0.52	1.78
Brazil	1.73	(-) 0.91	0.82
Chile	3.36	0.04	3.40
Colombia	2.28	(-) 0.18	2.09
Dominican Republic	2.98	0.39	3.37
Ecuador	1.15	0.13	1.28
Mexico	1.53	(-) 0.43	1.09
Peru	2.49	0.22	2.72

Source: Fernández-Arias and Fernández-Arias (2021); information for Honduras is not available.

The striking result is that, even for countries with positive **TFP** growth, the rate is below 1 percent, which is a value commonly observed in other countries. For the region as a whole over the same period, the average growth rates of **TFP** and per capita factor accumulation were -0.08 percent and 1.85 percent, respectively, yielding a per capita **GDP** growth rate of 1.77 percent—a result that, as Fernández-Arias and Fernández-Arias (2021) show, makes Latin America the world's growth laggard.[55]

[55] A complementary perspective is to consider the evolution of the productivity gap. In 1990, the productivity gap in South Korea vis-à-vis the United States was 35 percent; by 2018, it was 65 percent. In contrast, the simple average of the productivity gap of the nine countries considered here was 20 percent in 1990 and 26 percent in 2018; see Alfaro and Kanczuk (2020).

The social and economic costs of stagnant **TFP** are very large. If, for the region as a whole, **TFP** growth had been just 1 percent over the period considered, rather than -0.08 percent, average per capita income would have been 32 percent higher in 2017—a significant difference indeed!

Undoubtedly, lack of **TFP** growth has been the Achilles' heel of the region's growth strategy, a conclusion buttressed by the observation that after 1990, factor accumulation in Latin America was faster than in other regions of the world except East Asia (Fernández-Arias & Fernández-Arias, 2021).

Why was **TFP** growth so low and in some cases negative? This a difficult question and we do not claim to have a complete answer; clearly, country-specific factors play a central role here. However, to provide some region-wide insights, relation (11) presents a simple accounting decomposition over two periods, t and t + 1:

$$\mathbf{TFP}_t = a_{1t}TFP_{1t} + a_{2t}TFP_{2t} \quad ; \quad a_{1t} + a_{2t} = 1$$

(11)

$$\mathbf{TFP}_{t+1} = 0 + a_{2t+1}TFP_{2t+1} + a_{3t+1}TFP_{3t+1.} \quad ; \quad a_{2t+1} + a_{3t+1} = 1$$

In the first period, there are two firms, each with productivity TFP_{it} and shares a_{it} of resources $(I.H, K_{firm})_t$. In the second there are also two firms. However, between periods, firm 1 exits the market and firm 3 enters it. In the second period, firms have shares a_{it+1} of resources $(I.H, K_{firm})_{t+1}$ and productivities TPF_{it+1}.

Ideally, three outcomes are expected from relation (11):

- $TFP_{1t} < TFP_{2t}$, so that the low-productivity firm exits the market;
- $TFP_{2t+1} > TFP_{2t}$, so that the productivity of the surviving firm increases; and,
- $TFP_{3t+1} > TFP_{2t+1}$, so that the entering firm has higher productivity than the surviving one.

These outcomes describe standard Schumpeterian firm dynamics: on average, low-productivity firms exit, surviving firms become more productive, and entering firms are better than incumbents. Clearly, under these conditions, $\mathbf{TFP}_{t+1} > \mathbf{TFP}_t$.

Policies in M_{FTI}, M_{PDP} and M_{COM} in the first pillar of the region's growth strategy were all geared towards this fundamental Schumpeterian process of creative destruction. Foreign trade and investment policies would wean out low-productivity firms; competition policies would remove barriers to entry and impede inefficiency-inducing rents; and productive development policies would help surviving firms to become more productive and facilitate the entry of new higher-productivity ones. Policies in the second pillar would be the complement, increasing the supply of workers with more education and abilities, and improving the availability and quality of infrastructure.

To shed some light on why this did not happen, it is useful to write two alternative but equivalent expressions of **TFP**:

$$(12) \quad \mathbf{TFP} = [\, a_1TFP_1 + ... + a_{ni}TFP_{ni} \,] + [\, a_{ni+1}TFP_{ni+1} + ... + a_{(ni+nf)}TPF_{(ni+nf)} \,]$$

$$(13) \quad \mathbf{TFP} = [\, a_1TFP_1 + ... + a_{nT}TFP_{nT} \,] + [\, a_{nT+1}TFP_{nT+1} + ... + a_{(nT+nNT)}TPF_{(nT+nNT)} \,]$$

The first is the same as relation (7), repeated here for convenience. Firms are divided between informal and formal, with ni and nf, respectively, the number of each, and n (= ni + nf) the total number of firms. The second expression divides firms by their exposure to international competition, with nT firms producing tradeable goods, nNT producing non-tradeable goods, and n (= nT + nNT) the total number of firms. We emphasize that the firms in (12) and (13) are the same, in the first case classified by their contractual structure with their workers, and in the second by their degree of exposure to international competition.[56]

From relation (12), we can obtain the productivity distribution of all firms. By definition, those with high TFP_is would be in the right tail, and those with low TFP_is in the left one. From relation (13), we can also obtain the productivity distribution, which would be exactly the same as the first because they are the same firms. When we consider the productivity distribution in the first case, we would find that informal firms are overrepresented in the left tail, and formal ones in the right tail. Similarly, when we consider the second case, we would find that firms producing tradeable goods are overrepresented in the right tail, and those producing non-tradeable goods in the left one. This is just another way of saying that, on average, the TFP_i of firms producing tradeable goods is higher than that of those producing non-tradeable goods; similar results are found for formal firms compared to informal ones.

But independently of how firms are labelled, both distributions would have the same mean and the same measures of dispersion (standard deviation, difference between the TFP_i of firms in the 75th and 25th percentile, or the 90th and 10th). These measures reflect the extent of misallocation of resources in the economy. If it is low, differences in TFP_i between firms are not too large, and reallocating resources from firms with low TFP_is to firms with high TFP_is would increase **TFP**, but not by much. On the other hand, if it is high, reallocating resources from low to high productivity firms makes a significant difference for **TFP**.

The region's growth strategy looks at **TFP** through the lens of relation (13). Because firms producing tradeable goods are exposed to international competition, they need to continuously increase their productivity to survive. Exporting firms play a particularly important role because they can achieve economies of scale and are exposed to the newest technologies in the world, facilitating their adoption. Moreover, through backward linkages, productivity gains can be extended to domestic firms that provide inputs to exporting firms, including some that produce non-tradeable goods, particularly if solid productive development policies in M_{PDP} are in place; and if they are complemented by policies in M_{COM} to avoid collusive rent-seeking, and policies in M_{ENF} to improve the contracting environment. All this contributes to raise the a_is and TFP_is of exporting firms, or of firms that supply the domestic market but compete with foreign ones, so that over time, resources are reallocated to the more productive firms in the economy.

Importantly, at any point in time there can be misallocation in the sense that firms with low TFP_is attract more resources than they should, but this is secondary. In fact, misallocation could be interpreted as a sign of progress if some firms are pushing the right tail of the productivity distribution further out. What really matters is a vigorous process of structural

[56] Importantly, there is no one-to-one mapping between informal firms and firms producing non-tradeable goods, nor, therefore, between formal ones and those producing tradeable goods. There are formal firms in services and commerce, usually considered non-tradeable, and informal firms in manufacturing, usually considered tradeable; Levy (2018) documents this for the case of Mexico.

change or structural transformation that continuously increases the mean of the productivity distribution, underpinned by Schumpeterian firm dynamics. Firm churning would then reflect a virtuous process of exit, growth and entry and job churning would be its complement, as workers continuously move from low- to high-productivity occupations.

In turn, firm and job churning would be reflected in relation (12), the mirror image of (13). Overtime, the a_is of many informal firms would converge to zero. Firms that deploy technologies with very low TFP_is would gradually disappear, to be replaced by firms that use more modern and sophisticated processes, capable of constant technology upgrading and innovation. Many one-person firms would disappear as well, as their self-employed worker-owners found better earnings opportunities in larger, more productive firms with better growth prospects. New capital investments and net job creation would occur in formal firms, which would in turn provide workers with more productive jobs; put differently, formalization would result naturally from the region's growth strategy. Very importantly, this formalization would not be induced by formalization programs based on special regimes that could perhaps give workers improved social protection while keeping incentives for firm behavior basically intact; it would result from a genuine process of structural transformation, not tricks.

Over the last two decades, events in Latin America have by and large not followed this story line. To reiterate, **TFP** growth was either very slow or negative. Of course, this was not true of every country and some, like the Dominican Republic and Peru, did well relative to the rest of the region (although not so relative to countries in East Asia). But for the region as a whole, it was true.

Many reasons underpin this unfortunate outcome. In some countries, maybe the combination of policies in M(.) and T(.) was ill-designed; and in others it was perhaps well designed but implemented with insufficient forcefulness. In parallel, investments in infrastructure were probably insufficient everywhere, or projects were badly designed or executed. Maybe improvements in the coverage and quality of education were not as substantive as they needed to be. We do not make a judgement on these issues here.

3.8.3. Interphase between the region's growth and social protection strategies

But there is another reason behind low or negative **TFP** growth that, at least in some countries, has played a prominent role: the undesired consequences of overlapping and often contradictory social protection policies. To see this, consider that after the 1980s debt crisis, while the region was implementing its growth strategy, it was also implementing its social protection strategy which, at a high level of abstraction, can be depicted as:

(14) { reforms to and new programs in L(.) + T(.) }

result in { more insurance, less poverty, lower inequality }

As it happened, this strategy left the basic structure of the (L_{CSI} + L_{JS} + L_{MW}) regime pretty much intact, despite some reforms to the *modus operandi* of individual programs in L_{CSI}, particularly pensions. Importantly, while firms were being exposed to international competition, or being

asked to take more risks by adopting newer technologies, the legal framework governing their most important input—labor— did not change in any meaningful way. Rather, countries expanded or created non-contributory health and pension programs, sometimes raised minimum wages, fostered small firms through microcredits and special tax regimes, and introduced or expanded targeted transfers, sometimes in the form of CCTs and sometimes in-kind.

The point to highlight now is that the social protection strategy undermined the growth strategy, and sections 3.5 through 3.7 of this chapter can be seen in this light:

- while the growth strategy wanted productive formal firms to grow, the social protection strategy taxed them, and subsidized informal, unproductive firms;
- while the growth strategy sought to channel more resources and increase the market share of formal firms, the social protection strategy fostered the allocation of resources towards informal firms and allowed them to take away market share from formal ones;
- while the growth strategy sought to create jobs in larger firms, the social protection strategy rewarded self-employment and micro firms;
- while the growth strategy fostered backward linkages from a dynamic export sector to the rest of the economy, the social protection strategy hindered them;
- while the growth strategy required flexibility in firm-worker relations as firms were exposed to more competition, the social protection strategy kept them rigid;
- while the growth strategy promoted investments in education, the social protection strategy lowered the returns to education and thus the incentives to invest in education, and limited workers' opportunities to obtain training and learn on the job; and,
- while the growth strategy required stronger rule of law to improve the functioning of credit markets and the contracting environment, the social protection strategy *de facto* allowed tacitly accepted social norms to replace compliance with laws in a significant share of economic activity.

In sum, the social protection strategy was thinning the right tail of the productivity distribution and expanding the left tail, the opposite of what the growth strategy was trying to do. The social protection strategy implicitly supported a process of destructive creation, as resources were pushed into low-productivity activities, in opposition to the Schumpeterian process of creative destruction at the heart of the growth strategy (Levy, 2018). It was as if Penelope in Homer's Odyssey unknitted during the night through the social protection strategy what the growth strategy knitted during the day.

We make two points. First, while other policies in $\mathbf{E}\{.\}$ were also responsible for the stagnation of **TFP**, social protection policies contributed to this outcome; second, at least in some countries, this contribution was substantial and perhaps dominant because the resources channeled to social protection are large as a share of GDP; in fact, much larger than those channeled to promote innovation and technology adoption, subsidize credits to exporting firms, invest in infrastructure, grant investment subsidies to individual firms and, more generally, finance the policies in M(.) and T(.) associated with the growth strategy.

Of course, the contribution of social protection policies to stagnant **TFP** varies across countries, for two reasons already discussed but worth recalling here: first because the combination of policies that operate through the labor market and those that operate independently of it differs across countries, and second, because the parameters of Tables 7, 8 and 10 differ.

Unfortunately, the data available for most countries in Latin America are insufficient to closely track firm dynamics and quantify the factors discussed here. Brazil and Mexico and, to a lesser extent, Colombia, are important exceptions. In the case of Brazil, Ulyssea (2020a) combines cross-sectional information on informal firms and a high-quality panel data on formal firms with matched employer-employee data for both types of firms (unique in Latin America). He then simulates firm dynamics over a ten-year period and finds that firms grow 20 percent, compared to 50 percent in the absence of informality. Further, because informality allows inefficient firms to survive and take market share from formal firms, **TFP** is around 15 percent lower.

Mexico has detailed data on all formal and informal firms, allowing us to measure **TFP** according to relations (12) and (13). This country's growth strategy placed particular emphasis on policies in M_{TFI} and, judging by the behavior of exports, was very successful: these increased from 25 percent of GDP in 1996 to 35 percent in 2015 and diversified to the point where they now consist mostly of manufactured goods whose value exceeds the sum of manufacturing exports from the rest of Latin America combined. Many firms using world-class technologies were created in aeronautics, auto parts and electronics, among others. Through the lens of relation (13), **TFP** should have increased.

But this was not the case. **TFP** growth was in fact negative, and the little GDP growth that occurred resulted only from the accumulation of human and physical capital. Levy (2018) maps the 1998 and 2013 productivity distributions and finds that in both years, the left tail is by and large populated by informal firms and the right tail by formal ones. He then compares both distributions and finds that between these two periods, more resources were channeled to the left than the right tail. While investment in physical capital increased, the share allocated to informal firms also increased; while the labor force became more educated, informal firms attracted more workers than formal ones. The country's productive structure became more polarized, as judged by the fact that all measures of dispersion of the 2013 productivity distribution exceed those of the 1998 distribution. Despite the growth strategy, Mexico's firm dynamics over the period were dysfunctional: many productive firms exited, many unproductive ones entered and of the firms that survived, more transited from formal to informal than in the opposite direction.

Data for Colombia are not rich as Brazil's or Mexico's but are consistent with dysfunctional firm dynamics. Eslava et al. (2019) study firm dynamics over a period of 30 years. Unfortunately, their data only cover manufacturing and exclude firms with 10 or fewer workers, leaving out most informal firms. Nevertheless, they find that firm dynamics in Colombia are very different than in the United States, and that the "up or out" patterns found in the United States are much weaker in Colombia as a result of the survival of small unproductive plants and much weaker selection of productive new plants (or, in the language used here, that firm dynamics in Colombia are not Schumpeterian).

Considering the data limitations for other countries in the region, the report from CAF (2018) is consistent with the same result: persistency of small informal firms and large productivity differences between formal and informal firms; see also IDB (2010) and World Bank (2014).

We do not argue here that dysfunctional firm dynamics and the stagnation of **TFP** in the region result only from social protection policies. It is useful here to return to relation (9). All policies in $\mathbf{E}\{.\}$ matter, and some outside the realm of social protection are very important. Evidently, policies in the "realm of monetary and fiscal policy" to respond to volatility in $\mathbf{W}$, or to avoid large domestic imbalances, are central. Policies in the "realm of market conditions" can matter significantly, as the brief discussion of imperfect contract enforcement pointed

out. One can hardly expect Schumpeterian firm dynamics in contexts where the legal underpinnings of markets malfunction. In addition, insufficient or low-quality investments in infrastructure have clearly played a role. But, that said, we do argue two points:

- that *mutatis mutandis* across countries, social protection policies are partly responsible for the stagnation of **TFP**; and,
- that for too long, policymakers and analysts in the region have underestimated the importance of social protection for growth, wrongly considering the latter to be mostly an issue of factor accumulation and good macro and market policies.

3.8.4. Growth versus socially inclusive growth

Stagnant **TFP** does not imply absence of growth, of course. Countries can grow even if **TFP** does not, as Table 11 shows. They can do so because demography favors them, because there is a positive shock in $\mathbf{W}$, or because macro stabilization efforts result in an investment boom as uncertainty is reduced and medium-term prospects improve. Growth can also occur when large infrastructure investments remove bottlenecks that were hindering investments by other firms, as the expansion of the Panama Canal attests. Specific sectors of the economy can also be sources of growth like, *mutatis mutandis* in Latin America, tourism, agriculture, mining, energy or maquila exports. But this type of growth is hard to sustain because the conditions underpinning it eventually change or peter out. Sustained long-term growth requires positive **TFP** growth.

Socially inclusive growth is harder to achieve than growth, particularly in Latin America, which is characterized by long dated and deeply rooted inequities. This is so because two components are required at the same time. The first is effective policies to expand the coverage of insurance, reduce poverty and mitigate inequalities; this is the socially inclusive component of socially inclusive growth. The second is **TFP** growth; this is the growth component of socially inclusive growth. Sustained socially inclusive growth can only occur if both components are simultaneously present.

Socially inclusive growth has been absent in most countries in Latin America, in part because the socially inclusive component has been at odds with the growth component. This is not to say that there has not been growth, nor advances in social protection indicators, but it is to say that, with a few exceptions, growth has been slow or short-lived and many of the policies implemented to improve social protection have dragged down **TFP**.

Growth can be accelerated by strengthening the region's growth strategy. While maintaining macroeconomic stability, it can partly be done by investing more in infrastructure and improving the quality of the services derived from these investments as discussed, for instance, in IDB (2019, 2020b). But more infrastructure will not correct the factors listed in Tables 7, 9 and 10.[57] Policies in M(.) can also be strengthened. These policies can shift

[57] Levy (2018) compares the size and formal-informal composition of firms between the greater Mexico City area—with probably the best transport, telecommunications and related infrastructure in Mexico- and the rest of the country. He finds that in the greater Mexico City area 88.9 percent of all firms are informal, compared to 90.2 percent in the rest of the country. In turn, 88 percent of firms in the greater Mexico City area have up to five workers and 1.3 percent have fifty or more, compared to 92 percent and 0.7 percent in the rest of the country.

the mean of the productivity distribution to the right but will also increase its dispersion; they will hardly change the incentives in Tables 7, 9 and 10 that underpin the large number of low-productivity firms in the left tail. There will be clusters of firms or activities where productivity grows fast, but these firms and activities are unlikely to deliver the large-scale improvement in the labor market that the region needs, particularly since many are capital- or technology-intensive; the good jobs created will be measured, at best, in the hundreds of thousands, while the good jobs required are in the millions.

Yes, the region's growth strategy can be strengthened to accelerate growth and even result in increases in **TFP**, but by itself this will not deliver socially inclusive growth. Social inclusion is more than lowering an index of poverty or inequality or ensuring that everybody can satisfy their basic necessities. Although sometimes underappreciated, it also requires giving everybody an opportunity to display their talents, acquire new abilities and contribute to their country's process of structural transformation—it is about being included. It is about not being left on the side while the other half of the country progresses, even if one is no longer hungry.

At the end of the day, socially inclusive growth requires a radical transformation in the labor market so that, in addition to transfers, incomes increase because individuals' efforts are better rewarded, particularly for those at the bottom of the earnings distribution. In the specific context of Latin America, infrastructure investments by themselves cannot achieve this transformation. And the policies in M(.) and T(.) that are the heart of the region's growth strategy are also unlikely to achieve it because, unlike other regions of the world, they are confronting strong incentives against them.

This point is key. In Latin American discussions on growth, there is almost unanimous agreement on the importance of investments in education and infrastructure. There is an important debate on the relative weight that should be given to policies in M_{FTI} versus policies in M_{PDP} (but less so around M_{COM} and M_{ENF}). This debate in part derives from different interpretations of the reasons behind the success of South Korea and other countries in East Asia, and in part from different interpretations on the empirical evidence on various market failures that justify M_{PDP} policies on one hand, and the impacts on productivity of M_{FTI} policies on the other.

These are clearly important debates. But what is at times forgotten, or at least not given the importance that it merits, is that the elements in L(.) and T(.) associated with countries' social protection policies can make a large difference. When these elements result in highly dysfunctional labor markets, as is the case in the majority of countries in Latin America, the policy combination of more investment in physical and human capital and some agreed-upon mix of M_{FIT} and M_{PDP} is less effective. The radical transformation in the labor market required for socially inclusive growth either fails to take place; or happens at a pace that is simply too slow relative to societal expectations. In response, countries in the region strengthen social protection policies in L(.) and T(.) that tax formality and subsidize informality, which is precisely the opposite of what is required to achieve the desired transformation.

A few examples are illustrative. From 2003 to 2019, the Dominican Republic had one of the best growth performances in the region, doubling its real per capita GDP. The poverty rate fell from 32 percent to 21 percent. But informality increased slightly, from 52 percent to 55 percent, as did the share of workers that are self-employed or work in a firm with five or fewer workers, from 51 percent to 52 percent. Peru has competently managed its fiscal and monetary policies. Partly as a result of that, and partly as a result of a positive shock in $\mathbf{W}$ (metal prices), its real per capita GDP increased 84 percent between 2004 and 2019 —not as good as the Dominican Republic, but still among the region's best performances.

But productivity only increased by 0.2 percent annually, and in 2019, 78 percent of Peru's workers were still informally self-employed or working for millions of tiny firms.

Honduras grew less than Peru and the Dominican Republic in the last decade, but at a faster rate than the region's average. Between 2010 and 2019, its real per capita GDP increased 16 percent. But the poverty rate only fell from 60 percent to 59 percent and the informality rate stayed put at 80 percent. As mentioned, Mexico had the best performance in manufacturing exports, but **TFP** growth was negative. Its GDP growth was among the slowest in the region, and it is hard to argue that it was socially inclusive: labor informality fell only from 59 percent in 2005 to 56 percent in 2019 and extreme poverty from 11 percent in 2008 to 7.4 percent in 2018; and, over the same period, moderate poverty actually increased from 33.3 percent to 34.6 percent.

All these countries, large and small, had relatively solid macroeconomic management and some were among the region's top growth performers, but none experienced socially inclusive growth. They could all grow faster by increasing their investment rate and improving policies in T(.) and M(.) to raise **TFP**. But under their current social protection strategies, they would not experience socially inclusive growth; and if they strengthened their efforts to expand social insurance and lower inequality and poverty through more non-contributory programs or targeted transfers, they would likely slow down **TFP** growth.

With the benefit of hindsight, we consider that over the last two decades policymakers in charge of the growth strategy, absorbed by the challenges of responding to volatility in $\mathbf{W}$ and the complexities of policies in T(.) and M(.) to raise **TFP**, perhaps did not pay too much attention to the details of the social protection strategy, other than worrying that budgets constraints were respected and, perhaps, that pension systems generated sufficient long-term savings in domestic currency for investments in K_{inf}. Their assumption was that strong social protection policies would make sure that the fruits of faster growth would be shared more equally than in the past.

Again with the benefit of hindsight, we consider that policymakers in charge of the social protection strategy, focused as they were on the intricacies of policies in L(.) and T(.) to increase social insurance and lower poverty and inequality, perhaps did not pay too much attention to the details of the growth strategy, other than expecting it to generate sufficient tax revenues to pay for social programs. Their assumption was that strong growth policies would create more productive jobs, without perhaps full awareness that their efforts to improve social protection could obstruct that outcome.

If socially inclusive growth is the *desideratum*, it cannot be conceptualized as the superposition of two separate strategies, one for growth and another one for social protection. In Latin America, one cannot separate the mechanisms by which societies produce output from the institutions and policies that determine how it is distributed among the individuals who contribute to it, including the incentives associated with them.

If socially inclusive growth is the *desideratum*, priorities need to change. The first item on the agenda needs to be reform of social protection policies. This is not to say that efforts at macroeconomic stabilization should be abandoned, that we should not invest more in infrastructure, that efforts to improve education should be put to the side, or that policies in M(.) to raise **TFP** should be postponed until a better day—far from it. These efforts and policies need to be vigorously pursued. But it is to say that to deliver socially inclusive growth, and not just growth, the region's social protection architecture must first be reformed.

4. How Should Social Protection Be Structured?[58]

This chapter outlines how the region's social protection architecture can be structured to produce better social outcomes and facilitate **TFP** growth. We begin with four observations. First, the analysis is normative. We do not discuss issues of political economy, which are very country specific, although we do make some general remarks in the next chapter.

Second, we do not attempt to provide *the* model for social protection. Rather, our purpose is to discuss the features of a more effective social protection system—a broad vision, if we may say so. This vision is built around a central principle developed in this chapter: universality with respect to the relevant population.

Many suggestions are made to improve social protection: more transfers for the poor, higher minimum wages, a universal basic income, better non-contributory health or pension programs, and so on. We consider that a general principle is necessary to guide which policies are appropriate and who should finance them. Of course, a principle is not a rigid blueprint to be followed everywhere. The central principle that we advocate here allows for variants and combinations of policies, as well as specific parameter values for each, whose pros and cons depend on the characteristics of individual countries. What really matters is that the principle be understood as spanning three dimensions at the same time:

- all the population exposed to a given risk needs to be covered through the same program,
- the source of financing should be the same in each case, depending on the type of risk covered, and,
- when programs provide benefits in-kind, quality should be the same for all.

The third observation is that individual policies should not be designed or judged in isolation. What matters is that the social protection system, considered *in toto*, delivers the desired results. Designing and evaluating individual policies in isolation is insufficient for these purposes. A specific policy may have undesirable features, but when combined with other policies, produces the desired outcomes. A particular tax may be regressive, but generate sufficient revenues so that, when considered jointly with the transfer programs that it finances, the resulting package achieves the desired progressivity. Similarly, a specific labor regulation may on its own be considered insufficient to protect workers. However, when considered with other policies designed to protect them, that regulation may be desirable, particularly if it creates better opportunities for all—and more tax revenues for social programs—by increasing **TFP** and accelerating growth. The balance is key, and that balance may play out differently across countries, given their circumstances and political preferences.

Finally, as the discussion will make clear, restructuring social protection is inextricably linked with substantive changes to the tax and transfer system and, in most cases, a net increase in the tax burden. Taxation, however, is a complex issue whose discussion exceeds the scope of this essay. Rather than offering a careful analysis, we point out instead the areas where we think taxes and subsidies, *mutatis mutandis* across countries, could be adjusted.

[58] The discussion in this chapter is partly based on Levy (2008, 2019) and Antón et al. (2012).

We return to some of these observations in the last chapter, but it is important to keep them in mind as we present our arguments. That said, sections one through nine deal with different aspects of social insurance and section ten with complementary policies for the poor. Section eleven evaluates the role of a universal basic income. In turn, section twelve deals with taxation while section thirteen considers the implications of our proposals for income inequality. Section fourteen concludes.

4.1. Unbundling Risks and Insurance

Who should be insured against what, and who should pay? The general idea is to identify the different risks faced by individuals depending on what they do and their stage in their lifecycle, and to ensure that all those exposed to a given risk are covered and protected with the same policy. Depending on the nature of the risk, the appropriate policy may be in L(.) or in T(.), and this division in turn guides whether it should be financed through an earmarked wage-based contribution or some other source of revenue. Ideally, the resulting combination effectively insures the relevant population in each case, without heavily distorting occupational choices, firm sizes or the nature of contracts between firms and workers as occurs at present.

Graph 14 divides individuals by whether they participate in the labor market and, when they do, by their contractual status. The upper part considers all individuals and identifies two relevant contingencies to which they are exposed: illness and longevity.

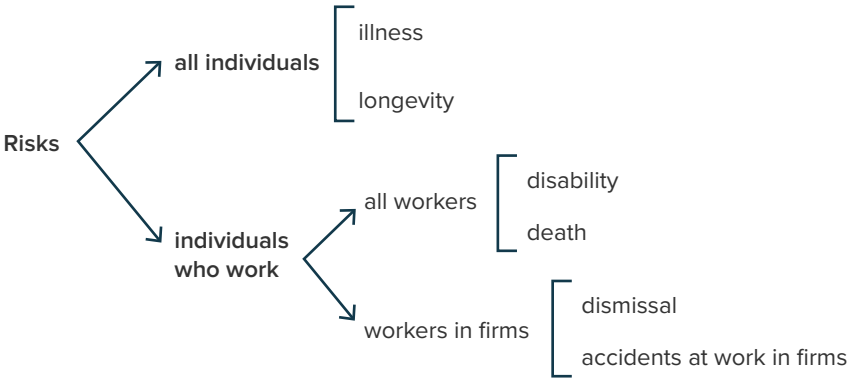
Consider illness first. Some individuals may never get sick, but others may suffer from illnesses of varying complexity and recurrence. This, of course, is not known *ex ante*; individuals of any age can be ill, and therefore all should be protected. Consider longevity next. All individuals will eventually die, although nobody knows when. This applies to all but matters specially for those participating in the labor market since they do not know whether their savings will suffice to support them in old age.

What are the policy objectives in each case? For illness, it is ensuring that individuals have access to health care when they need it; for longevity, that they have income during old age. Achieving the first objective requires a universal policy that covers individuals of all ages; achieving the second also needs a universal policy for the relevant population, in this case the elderly.

The lower part of Graph 14 focuses on individuals who work. The general contingency addressed here is that they may be impeded from doing so for some reason, with negative repercussions for them or their families.[59] Within that general contingency, we separate risks in two groups: first, those to which all workers are exposed, in particular death and disability, and second, those directly associated with a firm, losing the job because the firm goes bankrupt or dismisses the worker, and accidents at work, because firms fail to put in place safety measures (or because misfortunes happen even if they do). Addressing the first set of risks requires a universal policy covering all workers, including the self-employed; addressing the second requires policies whose relevant target population is workers in firms.

[59] This explains why disability and death are not included in the upper part of Graph 14, although of course individuals that do not work can also die or suffer a disability. The risk being covered is loss of income, and therefore the relevant target population is circumscribed to those working. That said, one should envision additional programs to help families with members with disabilities even if they do not work, as is currently the case in countries like Ecuador or Mexico; this would be part of a more encompassing social protection, discussed at the end of this chapter.

Graph 14: Individuals and Risks



Source: authors' elaboration.

4.2 Illness

Since all members of society can get sick, there is really no justification for providing health services through two different programs, one in L_{CSI} and one in L_{NCSI}, with all sorts of rules and requirements as to who is entitled to what and when. There is even less justification to do so when one considers that workers transition between formal and informal jobs, so that sometimes they (and their families) are covered by contributory health programs and sometimes by non-contributory ones, usually with different scope or quality.

The arguments made in chapter three all indicate that, from the point of view of social welfare and the point of view of **TFP**, it would be much better to provide health services of the same quality to all families, regardless of which of their members are working at any point in time or their transitions in and out of different jobs. This objective can only be achieved through a single transfer program funded from general revenues. This program, labelled here universal health insurance (UHI) would:

- significantly strengthen social inclusion since all members of society would always be entitled to the same publicly funded health services;
- ensure that, through taxation, cross-subsidies encompass all households, turning health services into a powerful redistributive instrument;
- maximize the degree of risk pooling since everybody would participate in the same program; and,
- delink access to health services from the vagaries of the labor market.

The first point is fundamental given the region's deep socioeconomic disparities. Discrimination and segmentation in access to health would end. UHI would perform the same equalizing function that publicly funded basic education has performed for many years.

The second and third points, jointly considered, are the underpinnings of the first. Unlike the current situation, high-income individuals would not escape from contributing to public health insurance because their earnings derive from profits or rents and not wages. Similarly, high-income self-employed workers would not escape either because they are not in a

relation of dependency, nor would those who are in a relation of dependency but deviate their contributions to separate funds. All households would contribute to finance UHI through taxation, and all would be entitled to receive services of the same quality. But because the amount of taxes paid by high-income households exceeds that paid by low-income ones, the result would be large subsidies from the former to the latter. UHI would be highly redistributive.

In parallel, by encompassing all members of society, health risks would be diversified as much as possible, increasing the efficacy of insurance (and thus lowering its economic costs). Further, a single program to which everybody is required to contribute would bypass the problems of adverse selection that occur when mandates are incomplete, or participation is voluntary.

The last point, finally, means that formality and informality would cease to be relevant categories for the purposes of access to publicly funded health services. Individuals could move in and out of the labor force, experience spells of unemployment or change from one job to another without any implications for their access to health; the labels attached to jobs (formal, informal, legal or illegal) would not matter. In parallel, occupational choices and contracts between firms and workers would be unencumbered by considerations of access to health.

Ignoring political economy considerations, the main obstacle to achieving UHI is fiscal. As noted in chapter three, with a few exceptions like Colombia, at present the quality of health services in L_{CSI} is higher than in L_{NCSI}. UHI should ensure that quality at least equals that provided by L_{CSI}. As a result, the gross fiscal cost of UHI would have two components: the cost of equalizing per capita health expenditures in L_{NCSI} to the level of L_{CSI}, and the cost of substituting wage-based contributions for health programs in L_{CSI} with other sources of revenue. The net cost, of course, would only be the first, as the second involves exchanging a wage-based tax for another one based on income, consumption, energy or something else. Separating gross and net costs highlights that, from a public finance perspective, the move towards UHI is partly a change in the composition of taxation and partly a net increase.

First-order approximations to gross fiscal costs for some countries are sketched out in section 4.12. They suggest that UHI is not a prohibitively expensive proposition, in practice out of reach. On the contrary, UHI is feasible, all the more so if attained gradually.

Importantly, the fact that health services are publicly funded does not imply that they have to be publicly provided. Many countries in the region already have large public infrastructure and personnel providing services (in health ministries or social security institutes), so in those countries the expectation is that provision would continue to be public. Others subcontract provision to private providers, and still others mix public and private provision depending on circumstances (level of care, complexity, regional availability). UHI is compatible with various mixes of private and public provision, a mix which should be pragmatically adapted to specific country circumstances. What is essential is that regardless of the mix, funding and regulation be public, and governments ensure effective access to all with the same quality, without any discrimination.

UHI is also compatible with private health insurance. Ideally, the range of illnesses covered and the quality of treatments offered by UHI would avoid catastrophic health expenditures or unnecessary suffering and deaths; all households should have the certainty that they will have support in the face of most illnesses. Ideally too, out-of-pocket expenditures would be substantially diminished from what they are today. That said, realistically, resource constraints will impose limits on some treatments, and complex ethical dilemmas as to what procedures are made available will inevitably arise (as is the case at present).

In this context, households should have the option of complementing UHI with private insurers, using their own resources. Similarly, firms should be able to negotiate with workers' complementary private coverage. Critically, however, these options should not dilute the obligation to contribute to UHI. This means that households' and firms' expenditures for private health insurance should not be deductible from personal or corporate income taxes. Ideally, the quality of services provided through UHI would gradually increase so that high-income households would use them, at least for some treatments. But even if this is not the case, or at least not in the short run, they should still contribute.

Finally, experience shows that the industrial organization of health provision is very complex. Even if the funding is there, providing quality services presents many challenges. One is to deliver the right balance of preventive and curative treatments. A second is to modify social habits with regards to diet, exercise and so on. Further challenges arise from population ageing and, as we have all learned, pandemics. The point here is that aside from funding, UHI also requires more effective organization of the health sector, including better regulation, more competition among suppliers of medicines and equipment, and stronger public health capabilities.[60] Once full coverage is achieved, UHI needs to be considered as a continuous process of learning, adaptation and improvement—a process guided by the twin goals of effective coverage and improved quality.

4.3 Old-Age Poverty and Longevity

All individuals face uncertainty as to when they will die. Some may have saved for retirement while they worked, but others may never have participated in the labor market or received a monetary remuneration for their efforts, a relevant consideration given the large number of non-remunerated workers in Latin America. Further, the deficiencies of the region's pension systems discussed in chapter three mean that many who save for a pension will not get one.

Two separate objectives need to be considered in this context: avoiding old-age poverty and helping income earners smooth consumption between work and retirement. The first applies to all the elderly regardless of their history of participation in the labor market, and the second to those that worked. In turn, two separate policies are required: one universal with reference to all the elderly and the second universal with reference to those that earned income before retirement.

4.3.1. Old-age poverty and universal non-contributory pensions

The simplest policy to avoid old-age poverty is a flat pension for all the elderly: a transfer program completely independent of individuals' prior history of participation in the labor market and income level, fully funded from general revenues. We henceforth call it a universal non-contributory pension.

A pension like this has many advantages. First, because everybody is entitled to one, it does

[60] The experience of OECD countries shows that indicators of health outcomes (life expectancy, morbidity, and so on) are not necessarily better in countries that spend more.

not affect individuals' decisions to contribute towards an additional pension when they work (unless it is very high). Second, it covers individuals who did work but never received monetary income. Third, it provides income to those who did save for a pension but were denied one due to the rules. Fourth, it avoids any exclusion errors associated with means testing. Fifth, it is highly redistributive: all the elderly receive the same amount, and all households contribute to its financing through taxes. But because higher-income households contribute more in absolute terms, low-income households come ahead in net terms.

As with UHI, the main obstacle is fiscal. Again, this varies across countries, depending on the size of the pension, demographics and the share of the elderly who already receive one. Section 4.12 provides rough estimates for some countries and shows that these pensions are within reach. Moreover, costs can be managed through a gradual expansion in the coverage and the generosity of the pension. What is central is to include all the elderly. This universal program, combined with UHI, aside from being a powerful expression of intergenerational solidarity, would go a long way to protecting them.

There is a large literature on these pensions. Evaluations show that they have a positive effect on the well-being of the elderly and their families.[61] That said, regulations need to be strengthened: how are they indexed to inflation, who sets the age at which they begin, and how this age is linked to changes in life expectancy (Bosch et al., 2013). Ideally, these regulations should be established in law, not only to avoid tinkering with them for short-run political gains, but also to turn these pensions into a social entitlement, as with basic education today and, hopefully, UHI in the future. There is also important scope to more fully integrate them with contributory pensions, as Chile has done (although in Chile it would be better if they were extended to all the elderly). All this needs to be adapted to each country's circumstances, of course. But given the past history of informal employment in the region, and the associated exclusion of the majority of workers from contributory pensions, a flat, universal non-contributory pension appears to be an essential component of the region's social protection system.

4.3.2. Consumption smoothing and universal contributory pensions

Contributory pension programs have traditionally been associated with workers in firms, where a contribution based on workers' wages links the size of the pension with earnings. Policy has proceeded under two assumptions: that workers have high discount rates or suffer from information problems such that they will not save enough for retirement on their own; and that once workers reach retirement, they need help managing the risks of longevity. Under these assumptions, regardless of whether contributory pensions are PAYG or defined contribution, at present these programs:

- require workers in firms to save for retirement; and,
- modulate the rhythm at which funds can be disposed once they retire.

There are many problems with this approach. First, some workers are excluded because

[61] See, for instance, Borrella-Mas et al. (2019) for Bolivia, Galiani et al. (2016) for Mexico, and Bando et al. (2020) for Peru.

they never associate with a firm, although they face the same discounting and information problems and longevity risks that workers in firms face. Second, others transition between formal and informal jobs, but are required to save only when formally employed. In their case, because the contribution density is low, the replacement rate—the pension as a share of the wage—is also low (compared to OECD countries), weakening the link between the amount of the pension and earnings. A third problem occurs in countries that establish minimum years of contribution to qualify for a pension or that prohibit pensions below the minimum wage. In these cases, workers who do not qualify lose all or part of their contributions or get them back in a lump sum at the time of retirement, *de facto* having to bear the risks of longevity.

In sum, self-employment and other contractual exclusions from L_{CSI} programs, formal-informal transitions and cumbersome rules result in contributory pensions that, as discussed in chapter three, protect Latin American workers very poorly while creating large problems in the labor market (without forgetting the high fiscal costs and often regressive properties of PAYG pensions when benefits exceed actuarial contributions).

To surpass these problems, it is essential to clearly identify the objective of these pensions. Is it to redistribute income, solve discounting and information problems or manage longevity risks? As discussed in chapter three, regardless of how L_{CSI} laws divide contributions between workers and firms, their incidence depends on the extent to which they were shifted back to workers. But note that even if, contrary to the empirical evidence, there was no shifting at all and firms paid for these pensions, their redistributive impact would be low because, as argued in that chapter, informal workers are left out and not all firm owners have higher incomes than formal workers.

Contributory pensions were created many decades ago when non-contributory ones were not significant. But we now know that the latter are substantially more redistributive than the former and that as long as they are universal, they do not generate problems in the labor market. Further, non-contributory pensions help to ameliorate longevity risks since the elderly always have some income until they pass away.

On this basis, we argue that contributory pensions should focus on consumption smoothing and, together with non-contributory ones, helping workers manage longevity risks. To achieve this, all workers should be required to:

- save for a contributory pension regardless of their labor status or income; and,
- participate in the same mechanism to pool the risks of longevity once they retire.

Requiring workers to save regardless of labor status is essential to ensure that contributory pensions protect all, not only those with long spells of formality. The only way to cover workers regardless of their transitions between formal and informal employment is to disassociate the obligation to save from the firm and transfer it to the worker; this way, no one is excluded.

There are two options for disbursing workers' savings when they retire. One is to let workers access the full amount accumulated the day they retire, like in Peru. In this case, workers deplete funds at whatever rhythm they prefer. The second, as in most countries, is to require them to buy annuities or, alternatively, to pay pensions from a common fund. The key difference is who bears the risk of longevity. While there are arguments for and against both options, on balance we think that the same discounting and information problems that justify requiring workers to save justify the need to protect them from longevity risks; the computation and perception problems are much the same (Benartzi & Thaler, 2007).

We therefore argue in favor of the second option. We also argue that to gather the largest possible pool of participants, there should be only one publicly regulated pension system, whether PAYG or defined contribution, with the same rules for all.

How can workers be forced to save regardless of whether they are self-employed or work in firms? The simplest mechanism to lower present consumption and increase future consumption is to tax current income and earmark the proceeds to increase future income. In turn, the simplest mechanism to achieve this is to:

- incorporate the requirement to save for a pension as part of the requirement to pay personal income taxes; and,
- divide revenues from personal income taxes into two components, one devoted to general government expenditures as at present and one earmarked for workers' own future pensions.[62]

Our proposal implies eliminating firms' contributions to workers' retirement pensions. Present practices as to how contributions for these pensions are divided between firms and workers vary in the region. For instance, in Chile and Peru, firms do not contribute and workers pay the full amount. Therefore, in these countries our proposal would only imply a relabeling: workers' personal income taxes would increase, while their contributions for retirement pensions would fall by the same amount. In other countries, firms and workers contribute. In Colombia for instance, firms pay three-fourths and workers one-fourth (12 percent and 4 percent of the wage, respectively). In these cases, depending on the level at which the contribution rate is set, workers could pay more than at present, but wages would adjust upward as the share of firms' contributions shifted back to workers disappeared (while formal employment expands).[63]

What about the self-employed? As noted, some countries like Argentina and Brazil require them to contribute to their pensions regardless of their income level; they already pay, some through special regimes that consolidate contributions to social insurance with personal income taxes. In Colombia, only those with earnings above one minimum wage are required to contribute, although in this case without any special regime. Thus, in these countries our proposal would imply a similar relabeling: the self-employed would no longer contribute to retirement pension programs but would pay higher personal income taxes with a share earmarked for their own pensions. The substantive change would be in countries that currently exempt the self-employed from contributing to L_{CSI} pensions, like the Dominican Republic, Ecuador, Honduras, Mexico and Peru.

Two points are relevant here. First, our proposal treats workers equally regardless of whether they are self-employed or work in firms or, more precisely, treats the same worker equally regardless of her formal-informal transitions. Second, our proposal needs to be considered jointly with two more that are described in more detail below: (i) negative income

[62] From a legal point of view, it would be essential to establish that the second component belongs to the worker. In fact, it would be better to relabel the personal income tax as the "personal income cum pension contribution tax" to underline that the contribution components is for pensions and could not be treated as general revenue.
[63] This is consistent with previous experience. In 2012, Colombia eliminated firms' contributions for health. Kugler et al. (2017) and Bernal et al. (2017) find that formal employment increased in response, particularly in small- and medium-sized firms.

taxes for low-income workers—in firms or self-employed—earmarked for their contributions to retirement pensions and (ii) contributions while workers are covered by unemployment insurance as part of this insurance.

While our proposal would have varying implications across countries, on the whole there would be five substantive changes:

- ensuring that contributory pensions cover all workers, not only those in firms;
- requiring that workers save for retirement regardless of formal-informal transitions;
- helping low-income and unemployed workers save;
- changing the legal nature of the contribution from an obligation of firms and workers to an obligation of workers that is incorporated into their personal income taxes; and,
- changing contributory retirement pensions from a program in L(.), to a tax-cum-transfer program in T(.).

Because everyone who pays income taxes would save for their pension, coverage would expand to include entrepreneurs, who in many Latin American countries are currently excluded from this requirement. This is important because there is no reason to think that they do not have the same information and computation problems than workers, and because not all of them have lower incomes than workers. (A third political economy reason is discussed in the next chapter). Universality applies to everyone who earns income, and the obligation to save would exist until they reach a certain age (say, 65 or 68). After that, income taxes could exclude the component earmarked for a retirement pension.

On the other hand, as with UHI, all individuals could also participate in complementary privately-run pension plans or, in the case of dependent workers, as part of their contract with a firm. Indeed, the government should help everyone do so, but without any preferential tax treatments.

Contributory retirement pensions would be directly proportional to individual savings, with no cross-subsides. Everyone who pays income taxes should have full certainty that their savings efforts will fully benefit them. There should also be no rules impeding individuals from getting a pension because they fail to comply with some minimum years of contribution or because the amounts accumulated generate pensions below the minimum wage. Further, there should also be no guarantees of a minimum contributory pension since the universal non-contributory pension is already a minimum pension guarantee.

In other words, we argue that all the cumbersome rules and requirements discussed in chapter three should be eliminated. All that is needed is an age at which individuals can begin accessing their contributory pension, which should be linked—although not necessarily equal—to the age at which they start to receive their non-contributory one. That said, individuals should be allowed to continue working and accumulating for their pension after that age. Indeed, rather than it being “the” retirement age, it would be the age at which individuals can access their pension savings. People could retire at any point after that age, or before if they want. Importantly, this would bypass discussions of retirement age to make the pension system financially viable.

Unlike UHI and universal non-contributory pensions, the main obstacle here is not fiscal; it is enforcement. At present, many informal workers do not pay personal income taxes because enforcement is weak; in fact, some are not even registered with the tax authorities. In this context, we make four points:

- as shown in chapter three, the informal earnings distribution overlaps with the formal one, sometimes quite substantially, implying there is ample room to tax informal workers;
- improving tax enforcement is something that countries in Latin America need to do, even ignoring savings for pensions;
- by linking contributory pensions with personal income taxes, incentives to comply may increase since workers would know that their pension will be proportional to their income taxes; and,
- enforcement is facilitated if, as proposed below, low-income workers face negative income taxes proportional to the contributions that they are required to make.

We highlight a very important implication of our proposal: if individuals save for their pension all the time that they work, contribution densities would be substantially higher than at present (although not 100 percent given temporary separations from the labor force). Higher densities could then allow for lower contribution rates without sacrificing the size of the pension. In turn, lower contribution rates would facilitate compliance. Put differently, countries could move from the current high contribution rate/low contribution density equilibrium to a low rate/high density one, with a significant increase in coverage.

4.3.3. Considering jointly non-contributory and contributory pensions

We argue for a more inclusive, simpler, fairer and transparent retirement pension system and for a more fiscally sustainable one. Expenditures for the universal non-contributory pension would be a flow payment expressly incorporated into the government's budget; there would be no contingent liabilities from contributory pensions. Subsidies for contributory pensions for low-income workers in the form of earmarked negative income taxes would be expressly considered in the personal income tax law. Everybody who works would save for their own contributory pension and, upon retirement, be entitled to it. Pension savings would rise, increasing funds for long-term investment projects. There would be maximum pooling of longevity risks and, considering the two pensions together, higher replacement rates and more effective consumption smoothing. [64]

Latin America has long debated the merits of PAYG versus defined contribution pension arrangements. As long as the contribution rate in the PAYG case is actuarially fair, our proposals are compatible with both. Each has advantages and disadvantages, which need to be assessed in specific country circumstances. A substantially more important issue, applicable almost everywhere, is that experience shows that under widespread informality, no arrangement can deliver the expected results. The region has experimented with parametric changes to PAYG or defined contribution arrangements, and many pension reforms in the 1990s tried to foster formality by introducing stronger connections between contributions and benefits, but on the whole, results have fallen short of expectations. With

[64] Replacement rates are usually measured only with respect to contributory pensions, but proper measurement needs to consider both. Replacement rates can also be increased by promoting additional voluntary savings through matching programs and information campaigns to change behavior towards savings. These measures are compatible with our proposals and could be an important complement. For a description of ongoing programs in the region see: https://www.iadb.org/en/labor-and-pensions/home-retirement-savings-laboratory

a few exceptions, coverage remains low and replacement rates high only for a minority of workers, usually higher-income ones, who have high contribution densities or benefit from government subsidies. This experience motivates our core proposals—that all individuals who work should save all the time and that contributory pensions should focus on consumption smoothing—which we believe to be significantly more important than whether they are implemented through PAYG or defined contribution arrangements.

The region's debate on contributory pensions has been complex because the objectives of these pensions have not always been clear, with results similar to what occurs when an eagle pursues two rabbits at the same time. Further, this debate has at times been delinked from the discussion of non-contributory pensions. What is needed is an assessment of whether the pension system, considering jointly contributory and non-contributory ones, achieves the desired objectives.

From this angle, we argue that in the context of the region's labor markets, the redistributive objective needs to be pursued through non-contributory pensions. That is why our proposals for contributory pensions are practically neutral from the point of view of redistribution (except for negative income taxes for low-income workers). Of course, a judgement may be made that the redistribution achieved through non-contributory pensions is insufficient, and this may be so. But in this case, these pensions or other redistributive programs in the social protection system should be strengthened.

On the other hand, if there is a universal non-contributory pension, could the contribution rate for the contributory pension not be zero? Why obligate workers to save if old-age poverty is averted? The answer depends partly on the size of the non-contributory pension and partly on the assessment of the behavioral and transactional impediments that workers face to save for retirement. For the reasons exposed, we think that the contribution rate should be positive. That said, we would argue for lower rates than currently observed in the region, which can exceed 20 percent. While it is difficult to set a fixed number without considering the generosity of the non-contributory pension, rates in the order of 5 to 10 percent would facilitate compliance and, in a context where contribution densities are higher, would result in a contributory pension that would be a reasonable complement to the non-contributory pension.

There is a trade-off in the composition of the total pension between its contributory and non-contributory components. A high non-contributory component reduces the need for a contributory component and implies more redistribution through general taxes. A low non-contributory component implies less redistribution and a stronger need for a contributory component through a higher surcharge on the personal income tax. This trade-off can be seen in a different light: how much of the total pension should be financed from general taxation with cross-subsidies and how much from individual contributions without cross-subsidies? Clearly, the answer depends on many parameters, but it is unlikely that it will be on one extreme or the other.

In sum, our suggestions for retirement pensions seek to simultaneously address the objectives of avoiding old-age poverty for all and helping those who worked before retirement to smooth their consumption over time. Through the two components, they also seek to manage longevity risks and increase replacement rates (which, we reiterate, should be calculated from the sum of both components). These suggestions need to be accompanied by parameter values: the generosity of the non-contributory pension, the contribution rate for the contributory one, the age at which workers can access each, and so on. These values matter immensely, and the right ones depend on countries' circumstances. But the broad

direction of change is clear: the retirement pension system should be universal, redistributive and foster long-term savings while seeking neutrality with regards to occupational choices, firm sizes and firm-worker relations.

4.4 Death and Disability

We now turn to the lower part of Graph 14. Although all individuals face death and disability risks, we focus only on those that work because, as mentioned, the issues that need to be addressed are hardships from foregone income. We proceed quickly since many of the arguments are similar to the ones just discussed. The key point is that all workers, not only those associated with a firm, can die unexpectedly or suffer a disability. As a result, these risks need to be addressed with a universal policy for all workers. If a worker suddenly dies, their relatives should receive a survivorship pension, and if a worker suffers a disability, she should receive a disability pension proportional to foregone earnings.

In most countries in the region, contribution rates for life and disability insurance are set at the actuarially fair level with no government subsidies or cross-subsidies between participants. This is as it should be. It is hard to see why survivorship or disability pensions should be redistributive. Contribution rates depend on benefit levels, but usually a rate of one to two percent of earnings suffices. Put differently, these protections are not expensive.

Universal coverage can be achieved by the same means described for the contributory component of retirement pensions. The actuarially fair contribution would be paid as part of personal income taxes. Importantly, the share devoted to retirement needs to be separated from the one devoted to life and disability. Further, the latter share should ideally be channeled to a common fund, maximizing the pooling of death and disability risks (and thus lowering insurance costs). As with contributory pensions, life and disability insurance would no longer be a contributory program in L(.) that only covers workers in firms, but a tax-cum-transfer program in T(.) that covers all workers. This change, however, would not be an impediment for firms to withhold payments from workers to minimize administration costs.

As with the contributory component of retirement pensions, the challenge for universality is not fiscal; it is enforcement. Importantly, the incentives to comply would be strengthened because workers would know that they and their families would be protected if death or disability suddenly impeded them from working. In fact, they may value this protection more than a retirement pension, which for some may be a highly discounted distant benefit.

Universality of life and disability insurance is essential because workers need to be protected regardless of transitions between jobs. Households who currently fall into poverty when the breadwinner dies or suffers a disability would be prevented from doing so; even those who are not close to the poverty line would be helped when they face hardships.

4.5 Accidents at Work

This and the next section focus on workers in firms, shown at the bottom of Graph 14. Here we consider the riskiness of the work environment. Miners, nurses, waiters and truck drivers face risks inherent to those activities. Accidents may happen. In extreme cases, they can result in workers dying; other accidents may permanently impair them from working. Worker compensation insurance protects them from these risks.

All countries in the region include this insurance as a program in L_{CSI}, in principle paid by firms (incidence issues aside). We think that this is as it should be since firms should be responsible for the safety of the work environment. It could be argued that self-employed workers also face risks when working, but these would be covered by disability or life insurance. The purpose here is to reflect the underlying risks of activities and modify firm behavior towards increased safety. Universality here is thus with respect to all workers in firms.

In some countries, contribution rates vary only by sector of activity; in others, they also vary across individual firms based on the frequency of accidents in each. The latter is better since it increases firms' incentives to invest in safety measures and avoids free-riding behavior. Contributions should be channeled to a single fund to maximize risk pooling. But because only workers in firms are covered, this fund should be separate from the one that manages contributions for life and disability insurance, which should include all workers.

4.6. Separation from Firms

Protecting workers when they separate from a firm is complex because issues of insurance, seniority and causes of and compensations for separation are simultaneously present. In Latin America, these issues are by and large tackled through job stability regulations in L_{JS}. However, a few countries like Argentina and Chile also include unemployment insurance, UI, as a program in L_{CSI}.

Restrictions on dismissal are understandable for two reasons. First, to the extent that access to health, life and disability insurance and saving for a pension are associated with a job with a firm, losing that job means losing coverage from those risks and a lower retirement pension, aside from losing current income. Second, job stability is associated with the acquisition of firm-specific skills, although it may also be considered valuable in its own right.

No wonder then that in Latin America, L_{JS} laws strongly obstruct and penalize the dismissal of workers. Generally, only misbehavior by workers is considered just cause of dismissal. Negative demand shocks or labor-saving technical changes are considered unjust causes, and when workers are unjustly dismissed, they can sue the firm and be entitled to large payments (that usually increase with tenure) and, in some countries, be reinstated in their job.

By and large, L_{JS} laws hurt the majority of workers. Because firms know that dismissing workers is very costly, they constrain their formal hiring. Further, when severance pay increases with tenure, firms' contingent liabilities increase with time. To minimize them, firms limit the number of workers contracted under these laws, offering the rest short-term contracts or contracts for pre-determined tasks (contractual forms which, *mutatis mutandis* across countries, are not covered by L_{JS} laws). They also subcontract some tasks to other firms or violate the law and hire informally. In parallel, firms only invest in training the subset of formal workers with whom they have a long-term relationship; the rest get little training and have fewer opportunities for on-the-job learning. These practices may lower firms' TFP_is, but they all yield higher expected profits than having their whole labor force under contracts that are covered by job stability regulations.

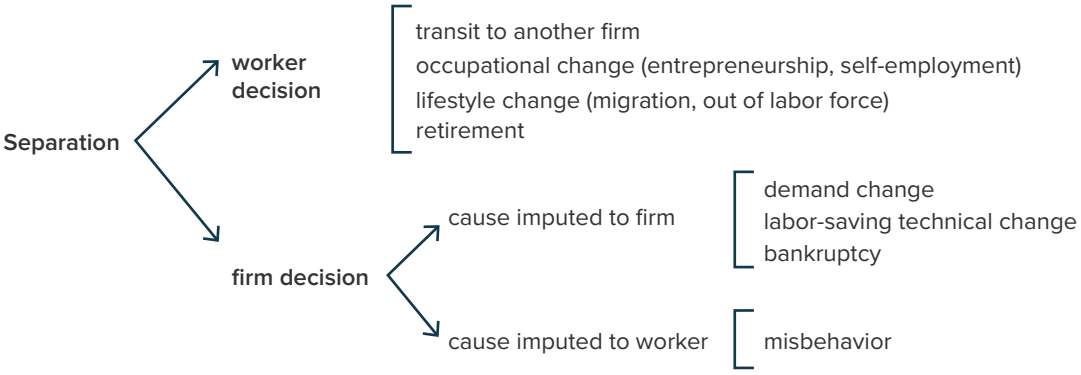
Yet even the lucky few with jobs covered by these regulations are not always protected. Firms may eventually have to dismiss them, and when this occurs, they may not receive the compensations that they are entitled to, often because firms face liquidity constraints (since they have to make large payments precisely when negative shocks occur). Further, suing firms

in labor tribunals when the cause of dismissal is considered unjust is costly and may have reputational costs for the worker. Even if workers win, they may collect their compensations long after the case is adjudicated, assuming firms do not fake bankruptcy or are not actually bankrupt. The incentives to settle out of court are large, with lower compensation than those stipulated in L_{JS} laws. Workers lose, firms lose and **TFP** suffers. The only clear winners are law firms and lawyers who specialize in litigating labor issues.

How can workers be better protected when they separate from a firm? We proceed in steps. We first identify the motives for firm-worker separation. Next, we identify the objectives that should be pursued in each case. Lastly, we propose a specific policy for each objective. Graph 15 helps with the first step.

The upper panel focuses on separations originated by workers. At present, depending on a country's laws, they may or not be entitled to a termination payment in these cases. Furthermore, even if they are, they may not get it immediately if the firm is liquidity constrained or facing a negative shock. The absence of an automatic, litigation-free mechanism for this payment may leave the worker without it or induce workers to force the firm to fire them so they can then sue for unjust dismissal rather than separate voluntarily. The lower panel considers separations originated by the firm, divided into those imputed to the firm and those (correctly or incorrectly) imputed by the firm to the worker (shirking and the like).

Graph 15: Separation between Firms and Workers



Source: authors' elaboration.

4.6.1. Cause free termination payments

Regardless of who originates the separation, workers should be entitled to a termination payment proportional to the time spent at the firm. To make it effective, we suggest that savings towards this payment should begin from the first day of hiring, channeling a contribution proportional to workers' wages to individual accounts legally owned by them.[65]

[65] Variants of this practice are already observed in Colombia, Honduras and Ecuador (labelled cesantías), among other countries, but are disconnected from the rest of protections to workers suggested in this section. The construction sector in Argentina also provides a useful example. Given the pattern of seasonal employment, employers contribute a fraction of the salary to a special bank account which workers can access when separation happens, while still being eligible for unemployment insurance. In Brazil there are also payments into workers' individual accounts with a similar purpose, under the Fondo de Garantia du Tiempo de Servicio (FGTS).

Resources would accumulate until the day of separation, when workers would have the right to automatically retrieve them.[66] This mechanism protects workers effectively since the termination payment is guaranteed regardless of the financial situation of the firm at the time of separation. Contributions would be made by firms through a program in L_{CSI} very similar to insurance for accidents at work, except that resources would not be pooled in a common fund, but channeled to workers' individual accounts. We label these payments "cause free" to underline that they would occur independently of the motives behind the separation.

4.6.2. Enhanced unemployment insurance and dismissal compensations

Turn to separations originated by the firm. The critical point here is that the policy combination depends on coverage against other risks, in particular health, life and disability insurance and retirement pensions. The ideal situation is one where there is UHI and, as discussed below, mechanisms are deployed to ensure that unemployed workers are covered by life and disability insurance and continue saving for their contributory pension. In this situation, the fact that a worker loses their job does not imply that they lose access to insurance, although of course it means losing access to current income. Under those circumstances, we argue that negative output shocks and labor-saving technical change should be considered just causes of dismissal.[67]

Of course, dismissed workers also need income, and UI should be one of the mechanisms to provide it. UI would be a program in L_{CSI}, as is the case already in Argentina and Chile, financed from contributions made by firms proportional to worker's wages. This is in contrast to Brazil, where UI is financed from other sources of revenue and induces opportunistic behavior by firms and workers. Contributions should be channeled into a common fund to maximize risk pooling among all participating firms and workers (which we consider better than self-insurance through individual accounts like in Chile).[68]

Unlike the current situation, however, we suggest that UI be made up of two parts:

- direct monetary payments to the worker; and,
- contributions for life and disability insurance and retirement pensions as if the worker were still employed, based on their wages prior to dismissal.

We call this enhanced unemployment insurance (EUI), and its purpose would be to ensure that unemployed workers receive some income, while maintaining the same insurance against death and disability and continuing to save for their contributory pension, as they

[66] Workers could leave resources in their accounts if they wish. If they get a job with another firm, new contributions would be added to accumulated ones. Workers could access their funds any time there is a separation and if they do not, when they retire, they could retrieve them in a lump-sum payment or use them to enhance their pension.
[67] Other causes of unjust dismissal would remain as they are, like gender or race discrimination. In these cases, workers would have the right to sue the firm and, if they win, receive monetary compensations or be reinstated in their job.
[68] We eschew a discussion of the duration of unemployment insurance and the generosity and time profile of payments. Usually, these last for six months and replace about 60 percent of the salary, but many combinations are possible, with flat or declining schedules or a lump sum payment up-front equivalent to the stipulated flow amount, which may provide incentives to shorten search times; see the discussion in Alaimo et. al (2015) and Bosch and Esteban-Pretel (2013).

did when last employed. Of course, contribution rates for EUI would reflect the cost of the additional protections.

Finally, we propose that firms make an additional one-time payment at the time of dismissal, equivalent to one or two months of wages, which we call dismissal compensation. This compensation has two purposes: (i) to introduce a cost that is fully internalized by the firm on top of contributions to EUI with the purpose of reducing free-riding behavior and avoiding frivolous dismissals, and (ii) to provide extra income to dismissed workers immediately in addition to the flow payments derived from EUI and the resources accumulated in their personal accounts for cause free termination payments.

The fact that this dismissal compensation is relatively small is important because it would not be too onerous for firms facing a negative shock, nor would it induce them to argue that they are dismissing a worker for causes imputed to said worker to avoid it. Few firms, if any, would want to litigate a payment of one or two months of wages in court. However, unlike the contributions for termination payments and EUI, which would be programs in L_{CSI}, this one-time payment would be a program in L_{JS}, paid by firms only if dismissal actually occurs—a contingent cost, not a flow cost.

Under these proposals, dismissed workers would have three sources of income immediately upon dismissal: two with full certainty—termination payments and EUI—and one with almost full certainty—the dismissal compensation.[69] Workers would also have access to health services and, during the duration of EUI, be covered by life and disability insurance and continue saving for their retirement pension. Put differently, unemployment would be a radically different state from what it is today in most countries in the Latin America.

For their part, firms would have more flexibility to adjust their workforce while contingent costs of hiring would be minimized and would not increase with tenure; uncertainty would be much reduced. The incentives to hire workers and develop long-term relations with them would be strengthened, while the incentives to switch between salaried and non-salaried contracts, simulate bankruptcies and so on, would be diminished. All this would result in less litigation and reduced labor rotation, more training, on-the-job learning and acquisition of firm-specific skills, jointly leading to steeper returns to experience and higher **TFP**.

Many countries in the region have so-called active labor market policies, a mix of labor training programs and information and intermediation services to shorten search times for unemployed workers and improve firm-worker matches; see Kluve (2016) and Escudero et al. (2019) for reviews. These policies have had mixed results, in part because in the absence of UI, workers cannot afford to search for better matches with firms. There are also not enough firms on the demand side of the market willing to hire formally. Our proposals would increase their effectiveness. Workers could afford to look for better matches because they would have income and insurance while they searched. In fact, EUI could be linked with labor training and job intermediation services, so that workers acquire additional skills while looking for a job.

---

[69] Blanchard and Tirole (2004) characterize the optimal design of protection against dismissal with two instruments, unemployment insurance and dismissal compensations. The addition of a third one here under the label of cause-free termination payments responds to institutional features of Latin America, where many firms, particularly small ones, can elude paying dismissal compensations due to a context of imperfect enforcement (including faking bankruptcies).

There are, of course, many variants of the proposals made above, as well as parameter values for the various payments. But the important point is to place the discussion of policies for worker-firm separation in the broader context of policies to protect workers against risks and to realize that as these policies evolve, job stability regulations should evolve as well.

4.7 Universal Social Insurance

We now discuss the overall protection against risks associated with the components of social insurance discussed before. For these purposes, the first column of Table 12 lists the relevant risks while the second summarizes how they are covered under the status quo. The last four columns summarize the discussion so far, focusing on the policy, coverage, funding and implications for redistribution and risk pooling.

Inspection of Table 12 highlights a fundamental feature of our proposals: universality with respect to the relevant population. This needs to be understood in three simultaneous dimensions: coverage of all those exposed to a given risk, coverage with the same policy and, in each case, funding from the same source of revenues.

Three principles underlie the mapping between policies, coverage and funding in Table 12:

- risks common to all citizens should be funded from general taxes;
- risks common to all workers, from workers' earnings; and,
- risks specific to workers in firms, from firm contributions based on workers' wages.

These principles represent a significant expansion of entitlements to social insurance, which at present are by and large restricted to workers under relations of dependency and subordination with firms or, when the self-employed are required to contribute, to a subset of protections. Access to health and a basic retirement pension would be a right of all citizens, completely delinked from their participation in the labor market. Access to life and disability insurance and a complementary retirement pension would be a right of all workers, as would be the obligation to contribute. Access to insurance against accidents at work and protections when workers separate from firms would be a right of all workers in firms and an obligation of the firms hiring them.

The principles result in a rebalancing of social insurance between policies in the "realm of entrepreneur-worker relations", L(.), and the "realm of taxes and transfers", T(.). They also result in a rebalancing of policies within L(.), between contributory programs in L_{CSI} and job stability regulations in L_{JS}. Most compensations when workers separate from firms would be pre-paid by flow payments in L_{CSI}, with smaller one-time payments in L_{JS} when dismissals actually occur. Protecting workers when they separate from firms would rely more on *ex ante* insurance and less on *ex post* lump sum disbursals, reducing uncertainty for both parties.

The proposals in Table 12 would substantially increase risk pooling across all individuals in society, strengthening the foundations of social insurance; they would, so to speak, make social insurance much more social than at present. Where pertinent, as in health, risk pooling would be associated with redistributive cross-subsidies; where pertinent, as in death, disability or work risks, no cross-subsidies would be involved. In each case, however, risk pooling would involve the largest possible group.

Table 12: Overall Coverage of Insurance

	Status Quo	Proposed	Coverage	Redistribution	Risk Pooling
Illness	Formal: L_{CSI} Informal: L_{NCSI} at times targeted by income level	T_{TRN}, no conditions. Funding: general revenues	All members of society, same services	Yes, across all taxpayers	Yes, through taxation
Longevity: old-age poverty	Formal: L_{CSI} (if workers qualify). Informal: L_{NCSI}, at times targeted by income level	T_{TRN}, no conditions. Funding: general revenues	All the elderly, flat pension	Yes, across all taxpayers	Yes, through taxation
Longevity: consumption-smoothing	Formal: L_{CSI} (if workers qualify) Informal: nothing	T_{PIT}, earmarked personal income taxes (negative for low income)	All workers, pension proportional to savings	Only for negative income taxes	Yes, through annuities or PAYG arrangements
Death and disability	Formal: L_{CSI} Informal: nothing	T_{PIT}, earmarked personal income taxes (negative for low income)	All workers, pension proportional to earnings	Only for negative income taxes	Yes, common fund
Accidents at the workplace	Formal: L_{CSI} Informal: nothing	L_{CSI}, firm contribution	All workers in firms, pension proportional to wages	Partly, from firms to workers (depending on incidence of L_{CSI})	Yes, common fund
Firm-worker separation	Formal: L_{CSI} + L_{JS}: sometimes UI + severance pay; narrow just causes, reinstatement Informal: nothing	L_{CSI}, termination payments, firm contribution L_{CSI}, EUI, firm contribution $L_{JS,}$ dismissal compensation; broader just causes, firm payment if event occurs	All workers in firms: Termination payments in all cases + EUI and dismissal compensation when originated by firm	Partly, from firms to workers for termination payments and EUI (depending on incidence of L_{CSI}) + Yes, dismissal compensation	Only EUI, common fund

Source: authors' elaboration. UI = unemployment insurance (only income); EUI = enhanced unemployment insurance (income + contributions for life and disability insurance, and savings for retirement).

We argue that when all components summarized in Table 12 are considered *in toto*, social insurance would be substantially better than at present because components derive from a crisper mapping between objectives and policies and because they complement and are compatible with each other. Jointly, they would protect workers and households from idiosyncratic risks more effectively than today. Of course, much would depend on parameter values, quality of services and so on. In the ideal scenario, all six components would be present with reasonable levels of generosity and quality, providing effective insurance

and making a significant contribution to reducing inequalities. And while countries in Latin America are far from that scenario, Table 12 provides a vision of what they should aim for—a vision that evidently needs to be adapted to specific country circumstances.

We henceforth call this vision universal social insurance, or USI.

4.8 Universal Social Insurance and Productivity

Here we focus on the implications of USI for **TFP**. To do this, it is useful to recall the expressions in the last column of Table 7 that measure the taxes and subsidies on formal and informal employment under the *status quo*, for convenience repeated here as relation (15):

(15)

Tax (+)/subsidy (-) on formal employment:	$[(1 - v_{CSI}).L_{CSI} + (1\text{-}v_{JS}).L_{JS}] - T_{TRN}(?)$
Tax (+)/subsidy (-) on illegal informal employment:	$L_{ENF}.F - [v_{NCSI}.L_{NCSI} + T_{TRN}(?)]$
Tax (+)/subsidy (-) on legal informal employment:	$- [v_{NCSI}.L_{NCSI} + T_{TRN}(?)]$

where, as in Table 7, L_{ENF} is the probability that a firm hiring workers illegally is detected and fined F, and $T_{TRN}(?)$ are transfers that, depending on countries' regulations, may be targeted on workers' income level or status in the labor market. The first two lines refer to dependent workers, and the third to other types of workers in firms and the self-employed, relevant to countries where these workers are not subject to ($L_{CSI} + L_{JS} + L_{MW}$) laws.

We begin by noting that rebalancing social insurance between L(.) and T(.) results in smaller wage-based contributions (all nominally paid by firms). L_{CSI} would just include contributions for insurance for accidents at work, EUI and cause free termination payments, with the first two channeled into common funds and the last one into workers' individual accounts. The two largest components of L_{CSI} in most countries today, contributions for health and retirement pensions, would disappear.[70] As a result, L_{CSI} would fall in all countries, sometimes quite significantly. Equally important, because L_{CSI} would only provide direct benefits to workers, with no cross-subsidies, their value to them would increase. While not reaching 100 percent, v_{CSI} would be higher than today.

Next, note that job stability regulations in L_{JS} would, in an expected sense, be less onerous to firms because they would pre-pay most dismissal costs through L_{CSI} and because changes in demand or technology would be just causes of dismissal. In parallel, workers would value L_{JS} more because compensations when firms dismiss them would be almost automatic, reducing uncertainty, delays, transactions costs and litigation.

The combination of numerically smaller L_{CSI} and L_{JS} and higher v_{CSI} and v_{JS} is central:

- the implicit tax on formal employment would fall, potentially by a large margin.

[70] In some countries, like Brazil, Colombia and Mexico, L_{CSI} also includes contributions for benefits unrelated to risks like education, sports clubs, housing or early child development programs. These benefits involve cross-subsidies among formal workers and sometimes from formal to informal ones. Clearly, they need to be financed from other sources of revenue. L_{CSI} should just include contributions for risks related to firm-worker interactions, without cross-subsidies.

In parallel, note that non-contributory programs in L_{NCSI} would no longer exist. Because there would be no social protection benefits conditional on being informal, we also get a central result:

- the subsidy for informality would disappear.

Ignoring minimum wages (discussed further below), these changes would increase formal employment. Informal self-employment would fall partly because employment in firms would expand and partly because it would no longer be subsidized by L_{NCSI} programs. There would also be fewer informal workers associated with the micro firms that sometimes hide behind the "self-employed" (including non-remunerated workers in family firms). Occupational choices and the formal-informal composition of employment would reflect an equilibrium substantially less distorted by L_{CSI} and L_{JS} laws.

The nature of contracts between firms and workers would also change. Firms would have no incentives to switch between short- and long-term contracts, or between contracts with a relation of dependency or subordination and other type of contracts in order to elude job stability regulations. Workers and firms could decide on the best contractual arrangements between them without any implications for workers' access to insurance. Firms would have more flexibility, and workers more security.

It is useful to refer to relation (4), repeated here for convenience as (16), where the superscripts 0, 1 refer to the *status quo* and USI, respectively:

$$\text{(16a)} \quad \mathbf{TFP}^0 = a_1^0 TFP_1^0 + a_2^0 TFP_2^0 + \ldots\ldots\ldots\ldots + a_n^0 TFP_n^{0.} \; ; \quad \sum a_i^0 = 1$$

$$\text{(16b)} \quad \mathbf{TFP}^1 = a_1^1 TFP_1^1 + a_2^1 TFP_2^1 + \ldots\ldots + a_n^1 TFP_n^{1.} \; ; \quad \sum a_i^1 = 1$$

Since under USI, self-employment would fall and informal micro firms with very low TFP_is would disappear, $n^1 < n^0$, so that economic activity would be less dispersed than today; average firm size would increase. Concentration of resources would occur in more productive units.

The dynamics of the labor market would change as well. Longer-lived relations between firms and workers would reduce labor rotation and facilitate more investments in worker training and expanded opportunities for on-the-job learning. Firm dynamics would be less dysfunctional because the social insurance regime would no longer implicitly subsidize the entry and survival of small unproductive firms. Productive firms would have fewer incentives to stay small to evade L_{CSI} or L_{JS}, given that the implicit tax would be smaller. Because surviving firms would be larger and have longer-lived relations with their workers, they could more easily obtain productivity gains.

In sum, we would observe that $\mathbf{TFP}^1 > \mathbf{TFP}^0$ as a result of three complementary effects:

- in relation (16b) the a_is of informal unproductive firms would fall while those of formal productive ones would increase;
- the TFP_is of surviving firms would gradually increase; and,
- the entry of firms with very low TFP_is would no longer be subsidized.

Of course, the proposals in Table 12 do not imply that misallocation would disappear, nor that firm dynamics would be fully Schumpeterian. **TFP** growth would still be punished by

impediments derived from special tax regimes for small firms, by issues of imperfect contract enforcement that limit firms' access to credit, or by other market failures in M(.).[71] But the proposals do imply that the obstacles to higher **TFP** derived from social insurance policies would be substantially diminished.

4.9 Complementary Policies Against Poverty

Poverty alleviation policy in the region is presently designed in the context of the formal-informal segmentation of the labor market and the associated division of social insurance into contributory and non-contributory programs. Policy rests heavily on income transfers to the poor through generalized subsidies for energy and the like, exemptions to consumption taxes, and targeted programs (some in-kind and some in cash), the latter usually in the form of a CCT. As discussed in chapter three, the exact combination varies across countries, as well as the rules and conditions for qualifying for different programs.

On the whole, these policies reduce income poverty and expand access to health and pensions. However, some deepen the forces the segment the labor market, making it more difficult for poor workers to find formal, higher-productivity jobs. In parallel, because the scope and coverage of insurance through non-contributory programs is limited, not enough is done to keep near-poor informal workers from falling into poverty when they face a negative shock. The result is that, at present, poverty policies focus more on helping the poor endure poverty and not enough on helping them escape it or preventing others from falling into it.

USI provides an opportunity for a different approach, based on three components listed in order of importance:

- assurance that poor households have access to the same insurance as all other households;
- aligned incentives so that poor workers access formal jobs; and,
- complementary targeted transfers.

All in all, income transfers to the poor would increase, but their composition between in cash and in-kind, targeted and universal, and current and future consumption would change. Critically, they would be carried out in the context of a labor market that, as explained below, is more favorable to the poor.

4.9.1. Equal social insurance for poor and non-poor households

Begin with the first two rows of Table 12. UHI would reduce poor households' out-of-pocket expenditures for health and avoid potentially crippling expenses. These households would have more income to spend on other goods, and not have to liquidate assets, including those

[71] But note that the programs in Table 12 could improve firms' access to credit because firms' contingent liabilities derived from hiring workers formally would be substantially reduced. In case of bankruptcy or non-performance on loans, banks could claim a larger share of firms' assets since fewer assets would be pledged to cover obligations to workers, which in most countries have priority over claims by other parties.

used for production, when a member of the family falls seriously ill. The implied value of the additional income transfer can be substantial. Because most poor workers are informal, at present they receive the per capita transfer implied in non-contributory health programs; under UHI, they would receive the one implied by contributory programs.

Equally important, poor households would have the same right to health care as all other households, including access to the same clinics and hospitals. Ending the current segmentation in health provision would contribute decisively to social inclusion. In parallel, a universal non-contributory pension would directly increase cash resources in poor households, particularly in countries where these pensions are non-existent or have little coverage.

Consider next the third and fourth row of Table 12. Given their informal status, most poor workers today are not covered by life and disability insurance, nor do they save for a contributory pension. To revert this situation, we argue that all poor workers, including the self-employed, should:

- be required to register with the tax authorities and file personal income taxes;
- be exempt from any payments; and,
- conditional on filing, receive a subsidy from the government in the form of a negative income tax for their contributions for life and disability insurance, and savings for a contributory pension.

Self-employed poor workers are already required by law to file in many countries, although there is large-scale evasion. Similarly, laws already require firms that hire poor workers to file for them even if no taxes are withheld, although only formal firms comply. Further, most laws also exempt workers with low earnings from these taxes. Thus, the novelty of our proposal is a negative income tax earmarked for life and disability insurance and for savings for retirement.[72]

This proposal has three objectives. First, create a mechanism to transfer income to poor workers for a mix of future consumption (savings for a retirement pension) and insurance for present contingencies (death and disability risks) separate from a mechanism to transfer income for present consumption (as discussed below). Second, create incentives for self-employed poor workers to register with the tax authorities and file (although, again, they would not pay).[73] Third, create incentives for firms to hire poor workers formally. As with everything else, the proposal needs to be adapted to specific country circumstances, but the key point is that these income transfers would help poor workers participate in the same social insurance programs as all other workers. Inclusion is central.

[72] This would imply an incentive to work similar to the Earned Income Tax Credit in the United States; see Blank (2002). Under our proposal, however, the negative income tax would be earmarked for life and disability insurance and savings for pensions. It is also feasible to add a cash component to directly increase the monetary wage as in the EITC, further increasing incentives for poor workers' participation in formal employment.

[73] The arguments here are based on a key lesson from the plethora of impact evaluations of CCTs over the last two decades: poor households respond to incentives. Still, these incentives may not be strong enough to induce many to comply, and probably need to be complemented with stronger enforcement.

4.9.2. Improving labor markets for poor workers

Consider again relation (14). For many reasons, the implicit tax on formal employment is higher for the poor than the non-poor. In some countries, they face higher L_{CSI} contributions. But even if this is not so, their valuations v_{CSI} and v_{JS} are lower: they may live further away from the facilities where contributory health programs are provided; they may face more severe liquidity constraints and are thus less willing to save for a contributory pension or are less likely to comply with the requirements to obtain one because they have shorter spells of formal employment; they may not be able to afford the lawyers to sue firms when unfairly dismissed; and so on.

For many reasons, the subsidy to informal employment is higher as well: non-contributory pensions represent a higher proportion of their earnings, facilities for non-contributory health programs may be closer to where they live and so on. Further, poor workers are employed by smaller firms. Because these firms face lower probabilities of being fined if they evade, illegal informal employment is higher among poor workers than non-poor workers. This tax-cum-subsidy configuration is aggravated when transfer programs in $T_{TRN}(?)$ are conditioned on being poor and informal, so that poor workers lose benefits if they get a formal job or if their income increases.

Since poor workers are the most affected by the current tax-cum-subsidy configuration and other impediments to firm growth, they would be the main beneficiaries of the change implied by USI. Their employment conditions would probably change more than others: less self-employment or employment in micro firms and more in larger firms, more formality, longer-term contracts, more opportunities to learn on the job or receive labor training, and so on. In a word, USI would help poor workers get what they most need: higher-productivity jobs with the same protections against risks as other workers.

4.9.3. Complementary income transfers for the poor

Including the poor in social insurance and improving the conditions under which they participate in the labor market would go a long way to reduce poverty. Nevertheless, the poor need additional help for two reasons. First, depending on the age composition of households, the cash transfer from non-contributory pensions may not be significant for some. Second, improvements in the labor market may be slow. As a result, complementary income transfers in cash to support their current consumption are needed, although these transfers would play a less important role in poverty alleviation efforts than they do today.

Under the principle of universality with respect to the relevant population, all poor households should receive complementary income transfers. In many countries in Latin America, CCTs are the main programs that provide these, and we suggest that they continue, with conditions pragmatically adapted to country circumstances. In some cases, there is room to expand coverage to ensure that all poor households are included, regardless of whether they have children or not; this means, very importantly, that households with elderly members would receive the transfer from the CCT and the universal non-contributory pension.

There is also room to improve their functioning and enlarge their scope, perhaps by linking them to programs that foster early child development. Aside from transferring income for current consumption, CCTs should continue to promote investments in the human capital of the poor.

Very importantly, eligibility should not be conditioned, as it is today in some countries, on workers not having a formal job. Equally importantly, the size of the CCT transfer should be calibrated to consider three elements: the need to support current consumption, the value of the in-kind and negative income taxes that poor households would receive under our USI proposals and the need to minimize the implicit tax on poor workers when their earnings place them above $\mathbf{Y}^{cutoff}$. In some countries, this may imply lower transfers than today, particularly in those where transfers from CCT are a *de facto* substitute for benefits that they would receive under USI.

Altogether, under our proposals poor households would be entitled to:

- unconditional in-kind transfers for health through UHI;
- unconditional cash transfers through a universal non-contributory pension;
- transfers through negative income taxes earmarked for life and disability insurance and savings for retirement pensions; and,
- targeted cash transfers for current consumption through CCTs.

This transfer structure would support poor households across their lifecycle: through CCTs when they have children and youth; through negative income taxes when members participate in the labor market; through non-contributory pensions when members are elderly; and through UHI all the time.

The total value of income transfers to the poor would most likely increase relative to what they receive today, although of course much depends on parameter values. But with the exceptions of transfers through CCTs, they would be carried out through the same social insurance programs that protect everybody else and be sharply focused on helping poor workers get formal jobs. The change in the composition of transfers is very important because it would lower the importance of targeted cash transfers for current consumption and thus reduce the size of the implicit tax inevitably associated with any transfer scheme based on current income levels, facilitating the transition out of poverty (as discussed in the next section).

Finally, our proposal for USI would also help near-poor informal households avoid falling into poverty when they face negative shocks because they would be insured against some of them. Aggregate poverty would fall gradually; poor workers would more easily escape poverty because they would find more productive jobs and because the number of near poor falling back into poverty when there are negative shocks would fall. Poverty policy would be more balanced between better insurance, improved conditions for poor workers to escape poverty with their own efforts, and income transfers, while helping near-poor households to avoid poverty when circumstances are unfavorable.[74]

[74] Of course, poverty policy would be complemented by other social policies not discussed here, including those associated with early child development, education and access to housing, internet, and so on.

4.10 Universal Basic Income

This section discusses the role that a universal basic income (UBI) could play in the region in the context of the proposals made above. Because the expression UBI is used in different contexts, it is useful to begin by clarifying the population to which the word universal is meant to apply. Sometimes it is applied to all individuals in the country (or all adults), but sometimes only to a smaller group, for instance the poor. Here it means all individuals; as a result, we define UBI as a permanent and unconditional cash transfer of the same amount to every member of society or to every person over 18 or 21.

Of course, the reference population can be a group other than all individuals or all the poor. In that case, UBI would be a permanent income transfer of the same amount to all individuals in that group. There is no problem with any of these usages, as long as the reference population is made explicit. Because here UBI is defined with respect to all individuals, rather than using the expression "UBI for the poor," we use the expression "an income transfer for all the poor."

UBI has four advantages:

- it provides a minimum income in the face of any contingency, thus acting like insurance against events that cannot be anticipated;
- it avoids the implicit tax on earnings associated with targeted transfers;
- it avoids most of the complications and exclusions associated with targeted transfers (although some could persist, like residency requirements); and,
- unless it is very high, it does not affect labor-leisure choices.

On the other hand, its fiscal cost is likely very high because it covers all citizens, or all adults, although this also depends on the size of the transfer. A basic income is a subjective concept that accepts many quantifications; here we eschew this discussion. Rather, we focus on distinguishing two overlapping but different objectives of UBI: to provide insurance and to transfer income.

4.10.1. UBI as permanent insurance

Table 12 focused on risks that can be identified *ex ante* and policies that could be deployed to address them. Risks that cannot be anticipated were excluded. However, these risks need to be considered because they can materialize into systemic shocks that cause harm to a large number of households. A recent example, of course, is Covid-19. But aside from pandemics, societies can also experience systemic shocks from climate change or other hard-to-foresee phenomena.

Social insurance systems need to help households face these risks, and some have suggested that a UBI can serve this function (Banerjee et al., 2019). We argue here that this statement needs to be assessed in the context of the existing mechanisms to protect households against risks. The case for UBI is not the same in a country with no social insurance programs as in a country where there is universal coverage of health, retirement pensions and protections against death, disability and loss of employment. In the first case, UBI serves as a substitute for other forms of insurance, and in the second, as a complement, insuring only residual risks not already covered by other programs.

Unlike countries in the OECD, at present identifiable risks—those listed in Table 12—are erratically and inequitably covered by the region's social insurance systems. If these systems are left intact, UBI would have to protect households from identifiable risks, not only residual risks. But to cover health, life, disability and other risks, the income transfer would have to be quite large. This is extremely difficult from a fiscal point of view, but also unnecessary. When risks can be anticipated and have a known probability distribution, it is better to pool resources *ex ante* from the population exposed to those risks and then help a subset of this population *ex post* when contingencies materialize. To provide payments to all regardless of whether risks materialize or not defeats the purpose of insurance; it is, put differently, very inefficient risk management. UBI would only be the appropriate policy for protecting relevant populations against identifiable risks if there were no social insurance programs in place, and no possibilities to construct them.

But this is not the case in Latin America. All countries in the region have social insurance programs, although they are segmented, at times inequitable and have large incentive problems. The first priority should be to improve them along the lines of Table 12. Adding a UBI as an additional insurance program on top of the existing ones would be doubly costly because countries would now have to pay the cost of UBI while continuing to pay the cost of their present segmented systems, including the associated **TFP** losses. The region first needs to fix what is there, not ignore it and bypass it with yet another program.

4.10.2. UBI as contingent insurance

That said, events that cannot be anticipated need to be addressed. These events can be very different in nature in terms of their duration, intensity and the population affected. It is extremely difficult, however, to design a policy to provide protection against something that is unknown. A pandemic may affect everyone, but a hurricane or a drought may only affect a region of the country or those involved in some activities (fishing, agriculture). Even pandemics can differ in terms of their duration and gravity. Future climate-related or health-related shocks may be very different from the ones that the region has experienced in the past.

Therefore, it would be appropriate to respond rapidly and flexibly on a case-by-case basis. In some cases, a UBI may be the right response, although it is unlikely that the associated transfer would be permanent. In others, it may be part of the response, and in others, the response may not require it. The critical point is to have the institutional and operational capabilities to implement a UBI in case it is needed as part of the policy response.

The region's experience with Covid-19 highlights the importance of this. The complex nature of the shock necessitated income transfers, particularly to informal workers, although weak institutional capabilities limited the speed at which they could be delivered (with the exception of Brazil). For the case of formal workers, however, given the absence of unemployment insurance, the appropriate second-best response was a mix of wage subsidies and tax deferrals or reductions for firms accompanied by credit guarantees, conditional on not firing their workers (Levy, 2020). For this latter of group of workers, UBI was the third-best option. For them, preserving their job, including the value of their firm-specific human capital, was better, and society at large would also benefit from the survival of pre-existing employer-employee matches, given that these matches were being destroyed by a completely exogenous systemic, but transitory shock. Put differently, in the case of Covid-19, a restricted and transitory UBI was desirable—restricted in the sense that the universe would be limited to informal workers, and

transitory in the sense that the transfers cease after the shock from the pandemic dissipates. But this UBI certainly would not substitute for health insurance and—in the case of workers who perished from the disease—for a survivorship pension for their family members.

This line of reasoning suggests that countries should consider UBI as part of their social insurance systems, but under the form of a contingent UBI to be deployed in combination with other policies based on the type of shock. This contingent UBI would have a different footing compared to the policies considered in Table 12. First, the word universal would be defined with respect to the population affected by the shock, which could be all individuals or all adults in the country, but usually not. Second, it would not be a permanent entitlement. Third, the size and duration of the transfer would be determined on a case-by-case basis. And fourth, it would not be included as a permanent expenditure to be funded on a flow basis from a government's normal budget, but as a transitory expenditure funded from other sources depending on the nature of the shock (general contingency funds, foreign borrowing, special taxes).

Critically, however, this contingent UBI would rely on the operational capabilities of the social insurance programs listed in Table 12. Countries could rapidly identify the relevant population in each case, and the tailor-made UBI could be delivered quickly because in principle everybody would be registered in pre-existing social insurance programs and with the tax authorities.

All that said, a contingent UBI could face risks of political manipulation. Ideally, objective criteria should trigger it, but this is difficult to accomplish when the triggering event is unknown. Moreover, flexibility in terms of amounts and population covered is essential. Cruces and Gasparini (2009) argue in favor of pre-defined broad interventions in response to emergencies, and this would be a step in the right direction; that said, strong mechanisms are still needed to minimize risks of manipulation and ensure that this important instrument of social protection is not eroded by political abuse.

4.10.3. UBI versus targeted programs to transfer income to the poor

An alternative objective for UBI is to redistribute income towards poor households, justified on the grounds that current social insurance policies by and large exclude them. But a redistributive argument for UBI based on the failures of these policies is like recognizing that they are immutable and accepting that countries in Latin America will permanently suffer from their deficiencies. It is hard to argue that because current social insurance policies do not redistribute towards the poor, a new program should be created for that purpose, leaving these flawed policies intact. Again, if policies are not working, they should by fixed, not bypassed.

Still, even if social insurance policies are reformed along the lines suggested in Table 12, the parameter values for the various in-kind and in-cash transfers may result in insufficient redistribution towards the poor, particularly in the form of income freely disposable for current consumption. In response to this concern, we argued above that poor households should receive complementary transfers in cash.

When transferring income to the poor, should countries cover all households or only poor ones? The redistributive case for UBI rests on three elements:

- various factors make targeted transfers undesirable;
- these factors cannot be remedied; and,
- they are sufficiently serious to merit UBI despite its much higher fiscal cost.

Two factors argue against targeted transfers: exclusion errors and negative incentive effects. Exclusion errors can be a serious concern but, in our view, do not justify a UBI. They can be minimized by expanding the coverage of existing targeted transfers and improving targeting techniques, rather than taking the costlier route of a UBI. Assume, for instance, that poor households are found somewhere between the second and third decile of the income distribution. In that case, extending the coverage of targeted transfers to households between, say, the fourth and fifth decile would go a long way to minimizing exclusion errors (while allowing larger inclusion errors). The cost would still be half of a UBI. Put differently, a UBI is simply not a cost-effective mechanism to transfer income to the poor. There really is no need to give a transfer to households in the top four or six deciles of the income distribution to help households in the lower four.[75]

What about incentive effects? As discussed in chapter three, any targeted scheme will always mean that households just below the threshold that separates the poor from the non-poor will lose benefits if their income increases beyond that threshold, thus facing an implicit income tax that could lower work effort. The evidence of these effects from the region's CCTs is mixed. On one hand, Alzua et al. (2012), Banerjee et al. (2017), Skoufias and Maro (2008) and Skoufias et al. (2013) by and large find no negative effects on work effort or labor participation.

On the other hand, negative effects on female labor force participation have been found in Brazil's *Bolsa* Familia program (Medeiros et al. 2008). The case of Argentina and Uruguay also merit attention. In Argentina, Garganta et al. (2017) find that the universal child allowance program *Asignación Universal por Hijo* (AUH) has large negative effects on married women: a 25 percent fall in the probability of participating in the labor market. This finding is important because the size of the transfer is quite significant, equivalent to a 35 percent increase in total household income for the typical poor household. In parallel, Bérgolo and Cruces (2021) find that Uruguay's CCT *Asignaciones Familiares* has large disincentives to formal work, inducing about half of those leaving formality into informal employment and the other half into non-participation. In other words, incentive effects can be a serious issue when transfers are large or tied to labor status.

A UBI avoids negative incentive effects because transfers are independent of income levels or labor status. That said, the importance of these effect depends on the size and nature of the targeted transfer. If it is high relative to poor households' income, as is the case with the AUH in Argentina, the effects can be large, at least for some groups of workers. But if the targeted transfer is small and disassociated from labor status, as we have suggested above,

[75] The argument is not about saving money. In fact, the resources unnecessarily transferred to those in the upper deciles of the income distribution through a UBI could be used to help the poor in other ways, like improving water and sanitation, access to internet, early childhood development programs and better quality of schools. The point is that from the perspective of poverty alleviation, the opportunity costs of UBI are quite high. Cruces and Gasparini (2013) calculate that in Latin America the fiscal cost of a CCT-type targeted transfer program and a non-contributory pension for the elderly would be, on average, 1.7 percent of GDP if targeted to the poor, but 5.2 percent of GDP if universal—a difference of 3.5 percent of GDP which, if channeled to the poor through housing, water and sanitation and education programs, could represent a huge change in their living conditions.

the implicit tax on earnings will be low and poor households will gain from their additional efforts. Moreover, if the size of the transfer changes with the lifecycle of the family because the largest share is driven by school attendance of its younger members, as it does in some CCTs, incentive effects could be further minimized (see Levy, 2006).

One can think of the transfers implicit in UHI and a universal non-contributory pension as variants of a UBI, the first in-kind and the second in cash. The key point is that the poor would receive them too and would continue to receive them even if they are no longer poor. These transfers would have no disincentive effects. In this context, an additional cash transfer targeted to the poor would be a complement to, rather than substitute for, the UBI-like transfers associated with USI. In that context, any negative incentives effects from a targeted transfer would be reduced because the transfer should not be high; the larger share of the help provided to the poor would occur through the mechanism described previously.

Summing up, we argue that in the specific context of Latin America, countries should include contingent UBIs as part of their social insurance systems and develop the capabilities to implement them along the lines discussed above. We argue that from the point of view of redistribution towards the poor—again, in the specific context of the region—the case for a permanent UBI covering every member or adult in society is, to say the least, very weak indeed.[76]

4.11 Taxation and Social Protection

We now consider changes in the "realm of taxes and transfers", T(.), to finance social protection along the lines sketched before. Three objectives need to be reconciled:

- social protection programs must have a solid and permanent revenue base;
- the structure of taxes should strengthen the incentives towards formality associated with the proposals on social protection so that they jointly contribute to higher **TFP**; and,
- the overall balance between social protection and tax programs should be progressive.

We discuss the first two objectives in this section and the third one in the next.

The fiscal effort required to finance social protection along the lines suggested varies across countries. There are differences in coverage and resources allocated to existing programs and there would most likely be differences in the scope and generosity of new ones. On the other hand, many specific country circumstances need to be considered when discussing taxation: endowments of natural resources like minerals or oil, distribution of tax competencies between national and subnational governments, openness to capital flows and so on.

Given the multiplicity of situations, no attempt is made here to suggest specific increases in spending or tax changes to individual countries. Rather, the discussion in this section has two

[76] We emphasize the specific context of Latin America. There is an important debate in developed countries on UBI, but the issue motivating it is quite different: the risks that workers face from automation and robotics. In these countries, a UBI may be part of the policy response to those risks. On the other hand, there may be developing countries in other regions of the world with no social insurance programs, in which case the insurance argument for a UBI could be stronger.

purposes. First, it provides a rough estimate of the fiscal effort required and identify the taxes and subsidies that could finance it—an illustrative exercise, far from a careful assessment for any country. The second purpose is to consider the implications of our approach for **TFP** and growth.

4.11.1. Identifying the fiscal effort

Our proposals require raising additional tax revenues for:

- UHI;
- universal non-contributory pensions;
- income transfers for the poor through a CCT or similar program; and,
- negative personal income taxes for poor workers earmarked for life and disability insurance and savings for retirement pensions.

The other components of USI listed in Table 12 would be funded from wage-based contributions or firms' one-time payments, or workers' personal income taxes, requiring no new revenues.

Begin with UHI. The gross fiscal costs would vary across the region. For Brazil, there are in principle none, since it already has a unified publicly funded health system that provides services to all, although of course there is always room to improve quality allocating more resources.

But most countries in the region are not in that situation. For them, the gross fiscal costs consist of three items: (i) replacing contributions for health programs in L_{CSI} with general tax revenues; (ii) raising per capita spending in existing non-contributory health programs to the level of contributory programs so that the quality can be the same; and (iii) additional revenues to increase quality and remove capacity constraint beyond the level currently provided by contributory programs. In what follows we ignore this last item because, while certainly welcome, it is extremely difficult to quantify and is not indispensable to set the basis for UHI. What matters at this point is establishing a unified public health system wholly funded from general revenues that provides the same services that formal workers receive to all, even if the quality is still far from what is desirable.

The relative weight of the first two items varies across countries. In Colombia, the second is costless because per capita spending in contributory and non-contributory health programs is already the same. In addition, the government already funds part of the costs of contributory programs; as a result, the costs of UHI would consist only of the first item, around 1.5 percent of GDP. In Mexico, per capita spending in non-contributory health programs is close to 80 percent of the amount spent on contributory ones, so the costs of UHI would be around 1.3 percent of GDP. Something similar happens in Peru, where the difference in per capita spending between programs is smaller, about 10 percent; as a result, the costs of UHI would be 1.8 percent of GDP (higher than in Mexico because informal employment is larger). On the other hand, in Honduras and the Dominican Republic the gap in per capita spending is larger than in Peru or Mexico, so costs would be 2.4 and 2.7 percent of GDP respectively.

These are all very rough estimates, and many factors need to be considered to obtain more precise ones. For instance, unifying the provision of public health services, which is currently divided between social security institutes and health ministries in many countries,

would bring efficiency gains, as the use of physical infrastructure, equipment, personnel and procurement systems is optimized. On the other hand, as the average quality of publicly funded health services increases, usage would likely increase, requiring more resources. But the estimates suggest that UHI, inevitably with some restrictions on the treatments covered, can be reached with a reasonable increase in spending, particularly if carried out gradually.

Next, consider universal non-contributory pensions. Again, the starting point varies across countries. Mexico already has one with practically universal coverage, so there are few, if any, additional costs beyond the 0.6 percent of GDP currently spent on it (unless the generosity of the pension were to increase). Other countries also have these pensions, but the coverage and amounts are low and, as a result, few resources are spent on them: 0.1 percent of GDP in Colombia, 0.0008 percent in Honduras, 0.11 percent in Peru and 0.009 percent in the Dominican Republic. CAF (2020a) estimates the cost of extending countries' existing non-contributory pensions to all the elderly who currently lack one, with the same generosity. They find that the cost ranges from 0.02 percent of GDP in Bolivia to 0.51 percent in Paraguay, with a mean value of 0.24 percent.

The costs of a targeted cash transfer program that covers all poor households also vary across countries since, again, initial coverages differ. Argentina, Brazil and Mexico already have CCTs with wide coverage, spending between 0.4 and 0.5 percent of GDP. In these countries, it may be necessary to spend a bit more to expand coverage to minimize exclusion errors, although perhaps this could be achieved by improving administration; in any event, the amounts involved would not be significant. In the Dominican Republic, the situation is similar: it already spends 0.6 percent of GDP on its CCT and reaching full coverage would only cost 0.1 percent more.

But that is not true of other countries. For instance, Peru's CCT has low coverage and costs only 0.04 percent of GDP; reaching full coverage would cost an additional 0.3 percent. Colombia and Ecuador are intermediate cases. Their CCTs have broader coverage than Peru's, and currently absorb 0.2 and 0.3 percent of GDP respectively. For these countries, the additional cost to reach complete coverage would be on the order of 0.1 to 0.2 percent of GDP. In Honduras, on the other hand, the situation is more challenging since its CCT currently covers only 15 percent of poor households; as a result, reaching full coverage would cost around 0.9 percent of GDP (although in Honduras there is an additional large non-targeted poverty program that could be redirected).

It is much more difficult to provide rough estimates of the costs of negative income taxes for the poor because we have no reference point from existing programs and because many variables are involved. Costs depend on the contribution rates for life and disability insurance (usually between 1 to 2 percent of earnings) and for savings for retirement (usually between 10 to 12 percent, although our preference would be lower, say 8 percent). They also depend on the number of workers whose earnings are below the threshold established in the tax schedule to qualify and the shape of that schedule. If, for the sake of argument, we assume that labor earnings of the poor are 4 percent of GDP and contribution rates add up to 10 percent, the costs of negative income taxes could potentially be as high as 0.4 percent of GDP.[77] Again, these are gross costs as some countries like Brazil, Chile and Mexico already have wage subsidies for formal workers.

[77] As described in the text, the tax schedule would have an undesirable kink around the threshold separating poor from non-poor workers. To avoid this, negative income taxes would have to fade gradually as earnings get close to the threshold which, depending on the smoothness of the tax schedule, would actually reduce the fiscal cost.

This is a significant number, but in fact it would be excellent news if it was spent, as it would mean that all poor workers are employed by formal firms or, if self-employed, file their personal income taxes. It is almost impossible to think of money better spent. But it is clear that this number is a high upper bound; it would take some years for compliance rates to reach 100 percent, if ever. Actual spending would initially be quite low, unfortunately. Further, note that the largest share of these resources, although registered in the government's accounts as current spending similar to other social protection programs, would really be savings for retirement.

The balance of this discussion differs across countries. On the whole, however, the additional fiscal effort to finance social protection along the lines discussed would vary from very little to 3.5 percent of GDP.

Importantly, this effort would not have to be done in one year. In fact, in the case of UHI, which is the costliest component, it could not be done in one year even if the resources were there because it is unlikely that public health administrators would have the operational capabilities to rapidly absorb large additional resources. Spending could more easily be ramped up for non-contributory pensions and cash transfers for the poor, which would add up to less than 1 percent of GDP almost everywhere. But even in those cases, expansions can be phased in overtime, perhaps by first extending coverage to those currently excluded, and then progressively increasing the generosity of the transfers, particularly of non-contributory pension (which would in turn allow for lower contribution rates for contributory pensions).

In our view, a gradual increase in resources for social protection is amply justified given the shortcomings documented in chapter three, particularly if they are channeled to programs that, aside from expanding coverage and improving benefits, also correct the main inequities and incentive problems that currently characterize social protection in Latin America. In fact, as some of the factors derived from current social protection policies that depress **TFP** are gradually eliminated, costs as a share of GDP could be lower because growth would be faster.

4.11.2. Identifying sources of revenue

The average tax-to-GDP ratio in the region is 23.1 percent compared to 34.3 percent for countries in the OECD. With the exception of Cuba, all countries have a lower tax-to-GDP ratio than the OECD average. That said, there is considerable variation, with some close to the OECD average, like Brazil with 33.1 and Argentina with 28.8 percent; others are far below, like the Dominican Republic (13.2 percent), Mexico (16.1 percent) and Peru (16.4 percent). But even countries in the middle are significantly below the OECD average: Honduras (22.3 percent), Chile (21.1 percent), Ecuador (20.6 percent) and Colombia (19.4 percent).[78]

If the OECD is taken as the reference, these numbers imply that for most countries in the region it should be feasible to gradually raise revenues to fund our proposals. Moreover, note that the net increase in the tax burden would be less since part of the additional revenues would replace wage-based contributions for L_{CSI} health and pension programs. Further,

[78] See OECD et al. (2020). Country comparisons of tax/GDP ratios are not without difficulties, given differences in tax structures and accounting practices. The comparisons above refer to simple averages and point out, correctly, that the tax burden in the Region is significantly below the OECD. However, if country ratios are weighed by GDP, the difference between the Region and the OECD narrows, 25.1 versus 31.8%, but still leaves a gap of 6.7% of GDP.

resources can also be obtained by reducing generalized consumption subsidies, rather than from higher taxes. This is not to say that the additional effort is easy, but it is to say that it is something that, with varying speed, the region could accomplish.

There are relevant differences in the composition of taxes between the region and the OECD. On average, Latin American countries collect less from income taxes and more from value added taxes. OECD countries on average collect more from taxes on the environment and property. However, there are important differences within the region, precluding broad generalizations.

To gain more insights, we consider briefly individual taxes. Begin with those on personal income. Average revenues are 2.2 percent of GDP in the region, in contrast to 8.3 percent in the OECD—a notable difference. Another statistic complements this observation: in the OECD, 24 percent of all tax revenues derive from personal income tax, in contrast to 9.7 percent in Latin America. Even countries with the highest collections fall far below the OECD average: Mexico, 3.4 percent of GDP; Brazil, 2.9 percent; and Chile, 2.7 percent. The gap is larger for countries like Argentina, Peru and Colombia, which collect 2 percent, 1.6 percent and 1.4 percent respectively, and even larger for countries that collect very little, like the Dominican Republic (1.2 percent). Taking the OECD as reference, the potential for additional revenues from the personal income tax is significant everywhere.[79]

There are many reasons behind low collections. Some countries do not consolidate all sources of earnings for the base of the tax base (interest, rents, dividends and wages), or have special treatments for some. Others have large loopholes, like excluding earners from whole sectors of the economy (e.g., agriculture) or allow large deductions for private expenditures on education and health or exempt pensions. Yet another reason, less noted, has to do with special tax regimes for micro firms: sometimes they are not registered as a separate legal entity and their owners pay taxes in a special regime of the personal income tax, not the corporate income tax. This is, in effect, a separate, substantially more favorable personal income tax regime. Of course, weak enforcement also plays a role since a large number of informal workers evade even if, as shown in the previous chapter, their earnings are similar to those of formal ones.

Our proposals for negative income taxes for poor workers would increase the progressivity of the personal income tax, but more is needed. In a context where social insurance is universal, there is little justification for deducting private expenditures on health or contributions for private pension plans from the tax base. Nor is there much justification for exempting pensions, particularly when, as in Colombia, Ecuador and Mexico, they are subsidized and concentrated among higher-income individuals. There is also little justification for not consolidating all sources of income and applying the tax schedule over a single base. In parallel, the base can be broadened, eliminating legal exclusions and special treatments. Further, in some countries, marginal rates can be higher. These efforts, combined with more effective enforcement, would strengthen the contribution of this tax to finance social protection, in some countries significantly; in fact, they could fully cover the costs of our proposals.

[79] Country values for taxes in Latin America were obtained from the database of the Interamerican Development Bank and the Centro Interamericano de Administración Tributaria; see https://www.ciat.org/idb-ciat-revenue-collection-database/?lang=en

Next, turn to corporate income taxes. On average, they account for 3.5 percent of GDP in the region, slightly higher than the OECD average of 3.2 percent. In fact, the share of these taxes in overall tax revenues is higher in the region than in the OECD, 15.3 percent versus 9.3 percent. However, the variance matters: revenues are 4.9 percent of GDP in Colombia, 4.7 in Chile and 4.1 in Peru; at the low end are countries like the Dominican Republic with 2.2 percent. All in all, these numbers suggest that corporate taxes provide fewer opportunities for additional revenues than personal taxes, although that does not mean that at least in some countries they are absent.[80]

One issue that depresses revenues are special tax regimes for micro and small firms when they incorporate as a legal entity and are taxed as firms. As mentioned, the assumed ownership of these firms by low-income households suggests to many that these regimes play a redistributive role. The changes proposed to social protection call into question their need since redistribution towards low-income households would be more direct and effective through our proposals for social protection and personal income taxation. In this context, eliminating these regimes would increase revenues, partly by raising rates and partly by reducing arbitrage opportunities that erode the tax base (like dividing a larger firm into two legal entities). In addition, it would remove some of the disincentives to firm growth that depress **TFP**.

Other measures could also increase revenues from this tax. Transfer pricing practices by multinational or large domestic firms engaging in international trade that allow them to shift profits to lower tax jurisdictions should be sanctioned more effectively. In some cases, tax breaks for firms locating in some regions could be reconsidered. The same could be said of some regimes for accelerated depreciation of investments. These issues clearly vary from country to country, as does the space for increasing revenues. Attention needs to be paid to international tax treaties and international tax competition, avoiding firm migration to jurisdictions with lower corporate taxes. But changes along these lines would ensure that corporate taxation also finances the additional spending in social protection.

Consider now consumption taxes, which mainly take the form of VAT in Latin America. At present, average revenues from this tax are 5.8 percent of GDP, in contrast to 6.8 percent in OECD countries. Again, the variance is large. Some countries collect more than the OECD average, like Chile, with 8.5 percent of GDP, Argentina with 7.6 and Honduras with 7.2, suggesting limited space for additional revenues from this tax. But others are in the opposite situation, like Mexico, where revenues are only 3.9 percent of GDP or the Dominican Republic (4.6). Countries in an intermediate situation, like Peru with 6.6 percent of GDP, Ecuador with 6.1 and Colombia with 5.7, also have opportunities for more revenues.

Differences in VAT-to-GDP ratios derive from differences in rates and evasion, but also from the extent of exemptions for certain goods and services. Chapter three pointed out that attempts to redistribute through these exemptions are inefficient because, in absolute terms, fewer resources are transferred to lower- than higher-income households.[81] On the other

[80] In many countries there is some arbitrage between personal and corporate income taxes. This may explain why personal income taxes in the Region are so much smaller than in the OECD, while corporate ones slightly higher. That said, if the comparison is made adding both, the Region is still substantially below the OECD.

[81] Mexico provides an example. About 45 percent of the consumption basket is exempt from the VAT (either the final sale or also the intermediate inputs). But given the country's high inequality, between 35 to 50 percent of the revenues forgone are captured by households in the highest two deciles of the income distribution and only between 4 to 8 percent by those in the lowest two deciles (Dávila and Levy, 2003).

hand, many believe that eliminating exemptions would be regressive since exempt goods and services represent a higher share of low-income households' consumption basket (typically food, medicines and transportation). Recent evidence suggests that in a context of high informality this is actually not the case since a significant share of the consumption of the poor occurs in establishments that elude the VAT (Bachas et al., 2020).

Rather than assessing these viewpoints, we recall a point made before and to which we return in the next section: what matters is not the progressivity or regressivity of any individual tax on its own, but that of the whole array of taxes jointly with the social protection programs that they finance. In this context, in countries where VAT collections are low, partly as a result of exemptions, eliminating or substantially narrowing them could contribute considerably to finance social protection.

Another aspect of the VAT merits attention. In most of Latin America, it is applied under the credit method: firms selling intermediate inputs issue invoices including the VAT, buying firms deduct VAT paid on intermediate inputs so that consumers of final goods only pay tax on value added. However, most informal firms are not registered with the tax authorities, and when they are, they usually pay taxes under special regimes which may impede them from selling to formal firms because they cannot issue invoices. The interphase between these regimes and the VAT implies that informal firms can sell only to other informal firms or to final consumers, further eroding revenues from the VAT. In other words, eliminating exemptions and special regimes would have the double benefit of increasing revenues and removing obstacles to firm growth.

Turn next to other taxes and subsidies. On the tax side, there is space to increase revenues by taxing property and the environment. Taxes on property, which tend to be progressive, are low in the region: 0.8 percent of GDP versus 1.9 in the OECD. As before, differences between countries matter. For instance, Colombia collects 1.6 percent of GDP, close to the OECD average, but many countries collect very little: Honduras, 0.6 percent; Peru, 0.4; and Mexico and Paraguay, 0.3. Taxes on activities that hurt the environment are also higher on average in the OECD: 2.3 percent of GDP versus 1.1 percent in Latin America. In this case, raising them would help improve the environment, an increasingly relevant consideration even if social protection issues are ignored.

On the subsidy side, some countries in the region have costly generalized subsidies for energy (gasoline, electricity or gas). IDB (2017b) estimates that, on average, countries spent almost 2 percent of GDP a year on subsidies between 2008 and 2014, although there is wide variance. At the high end are Ecuador (6.4 percent) and Argentina (3 percent); in the middle, Mexico and Honduras (1.9 percent each), Colombia, (1.4 percent) and (Peru 1 percent); and at the low end, Brazil (0.2 percent) and Chile (0.1 percent).

Energy subsidies redistribute inefficiently, probably more so than exemptions to the VAT since expenditure shares of low and high-income households are quite similar or even higher for the latter (e.g., gasoline). Reducing these subsidies may be particularly relevant in countries like Argentina that already have high tax-to-GDP ratios.

Finally, consider tax enforcement. Accurate estimates of evasion in the region are hard to obtain since the border between elusion and evasion is sometimes gray. Foregone revenues from exemptions to income or value added taxes are hard to separate from outright evasion. Complex laws also sometimes make it difficult for taxpayers to fully comply, even if there is no intent to evade. The Economic Commission for Latin America and the Caribbean (ECLAC, 2020) presents a valuable discussion of these issues, pointing out large variance in evasion both between and within countries across different taxes. For the region as a whole, the average is 30 percent, which compares unfavorably with an OECD average of 11 percent.

Estimates of evasion of income taxes are scarcer, but the ones available suggest it exceeds evasion of the VAT. Evasion rates are in the order of 57 percent in the Dominican Republic, 49 percent in Argentina and 32 percent in Peru. In parallel, evasion rates of corporate taxes are 20 percent in Mexico, 31 percent in Chile, 49 percent in Argentina and 51 percent in Peru (ECLAC, 2020).

The balance of this discussion varies across the region. Considering all taxes and using the OECD as reference, there clearly is significant space to increase them in most countries, although these spaces are narrower in a few countries, making it necessary to reduce generalized subsidies (or adjust other expenditures). That said, we cannot assert that, on a country-by-country basis, extra revenues would match the needs for higher spending in social protection. Each country would have to evaluate parameter values for their social protection programs, including the time period in which universality would be reached and the options for raising additional revenues.

Levy (2019) provides an exercise for Mexico. The additional fiscal cost to reach UHI is estimated at 1.25 percent of GDP and for a universal non-contributory pension at 0.11 percent. An additional 0.2 percent is also contemplated to expand the country's CCT to ensure full coverage. All in all, the additional fiscal costs would be 1.56 percent of GDP. The exercise includes introducing unemployment insurance along the lines discussed above but financed from firm contributions. Despite this, contributions to social insurance would fall, as the components destined for health and retirement pensions are eliminated. To finance the fiscal cost, Levy proposes limiting special regimes in the VAT without changing the general rate (yielding 1.1 percent of GDP); reducing exemptions in the personal income tax for private health insurance and pension plans (0.12 percent of GDP); and reducing exemptions in the corporate income tax including some associated with special regimes for small firms (0.2 percent of GDP), again without increasing the general rate. This combination would leave a gap of 0.14 percent of GDP, which could be closed through lower evasion, based on the fact that the proposal promotes formality and facilitates tax administration.

Of course, many other combinations of benefits and tax sources are feasible. We use this exercise only to demonstrate that the change in social protection that we advocate is feasible. Each country, of course, faces different starting points, would perhaps set different benefit levels, and so on. The tax effort required would be different or present other trade-offs like reducing generalized subsidies or changing the composition of expenditures. Further, gradual implementation would facilitate the transition. That said, the discussion hopefully conveys that, *mutatis mutandis* across countries, it is possible to design a combination of higher taxes and reduced generalized subsidies to finance a shift towards universal social protection. This conclusion can be restated: fiscal considerations are not an insurmountable obstacle to achieve universal social insurance.

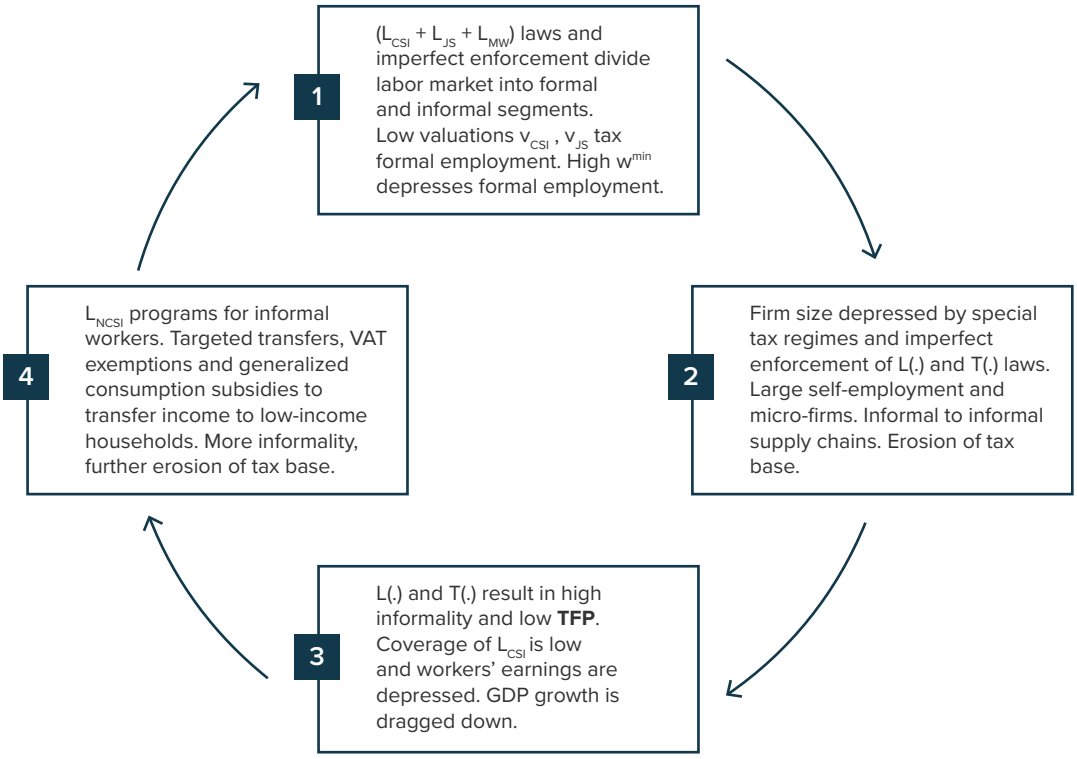
4.11.3. Taxation, informality and productivity

Discussions of taxation and social protection often focus on the role of the former to finance the latter; as long as revenues cover expenditures, tax policy is doing its part. In turn, as long as social protection policy expands coverage of insurance and transfers income to low-income households, it is doing its part. Throughout this essay we have argued that this is far from enough and that the incentives associated with social protection and tax programs are central to social protection and productivity outcomes.

Graph 16 is a stylized depiction of a vicious circle in which the region may be caught. It should be seen as a complement to Tables 7, 9 and 10 and Graph 12. It is not, of course, a description of any country, but an attempt to convey the idea that, *mutatis mutandis* across countries, at present there is a negative feedback loop between tax and social protection policies.

The point that we want to make here is central: proposals on taxation have to consider not only generating the revenues required to finance social protection, but the need to do so through a combination of taxes that increase formality, raise **TFP** and combat evasion. Not any taxes will do. Of course, the space to do so varies across countries given their current tax structures, but from the point of view of **TFP** two areas need special attention: special regimes for small firms and differentiated VAT rates. Special tax regimes may have important revenue costs, but even if they do not, they allow the survival of small unproductive firms that take away market share from more productive ones and obstruct small productive firms from growing. Differentiated VAT rates are also problematic, particularly when combined with special tax regimes, as backward linkages are weakened, informal supply chains established, and firm growth limited.

Graph 16: Vicious Circle between Taxation and Social Protection



Source: adapted from Levy (2019).

Special regimes and differentiated VAT rates are motivated by redistributive objectives. However, our proposals for social protection and income taxation allow these objectives to be reached more directly and effectively. These proposals would help revert the vicious circle depicted in Graph 16 and create a virtuous circle of more effective social protection, higher **TFP** and an expanded tax base.

4.11.4. Further observations on taxation and social protection

Considered jointly, the proposals described in the previous sections imply a considerable change in the financing of social protection, lowering the importance of wage-based contributions and raising that of other taxes. Some might interpret this change as letting firms off the hook, weakening the redistributive impact of social protection. It is exactly the opposite.

This issue is central to the debate on financing social protection, and we use Table 13 to elaborate on it. On the left side we list some of the views that sustain the current policy of strong reliance on wage-based taxes. We do not claim that any individual in particular holds these views; rather, we believe that they are consistent with current policies. On the right side we list the alternative views underpinning our proposals.

Table 13: Views on Taxation and Social Protection

Views underpinning current policies	Views underpinning our proposals
• All firm owners have higher incomes than workers.	• Family-owned micro firms and self-employment create a fuzzy border between firm owners and workers. Some firm owners have lower incomes than some workers. The vast majority of firms are very small.
• Firms absorb their contributions to L_{CSI} in the form of lower profits. Moreover, formal employment is invariant to contributions for L_{CSI} and regulations in L_{JS}.	• Firms' contributions to L_{CSI} are partly shifted back to workers in the form of lower wages. Workers' valuations of L_{CSI} and L_{JS} programs matter greatly for formal employment.
• The number, size and type of firms depends mostly on policies in M(.) and, to a lesser extent, T(.); policies in L(.) matter little.	• Occupational choices, firm sizes and the composition of employment are endogenous to $\mathbf{E}\{$ L(.), T(.), M(.) $\}$, with L(.) playing a prominent role.
• Formal workers have higher earnings than informal ones.	• Mean formal earnings are higher than informal ones, but the distributions overlap, sometimes considerably
• Informality and low **TFP** result from insufficient investments in physical and human capital; social protection policies in L(.) and T(.) do not matter much.	• Informality and low **TFP** are jointly determined by $\mathbf{E}\{.\}$, and social protection policies in L(.) and T(.) matter a lot.
THEREFORE: • To lower inequality, tax the formal sector and subsidize the informal one without worrying about **TFP**.	**THEREFORE:** • To lower inequality, tax high-income individuals and subsidize low-income ones, regardless of whether they are firm owners or workers. To raise **TFP**, tax informal activity and subsidize formality.

Source: authors' elaboration.

We consider that the empirical evidence presented in chapter three is more consistent with the column on the right, and highlight two substantive implications:

- in Latin America it is more efficient to tax high-income individuals directly rather than through firm contributions to L_{CSI} because the association between being a firm owner and being a high-income individual is weak, and because the vast majority of firms is very small; and,
- redistribution towards low-income informal workers should be motivated by the fact that their income is low, not by the fact that they are informal. Since subsidizing informal

activity to help low-income workers tends to perpetuate informality, it is better to delink subsidies to low-income workers from their status in the labor market.

Our proposals seek to tax high-income individuals without hurting the development of firms, and to help low-income individuals without subsidizing informality. In fact, we want to foster the development of firms and punish informality.

Of course, the impact of our proposal on income distribution and **TFP** depends on the taxes deployed to substitute for the wage-based firm contributions that we believe should be eliminated. That is why above we argued for greater progressivity for personal income taxes, reduction or elimination of special regimes for small firms, increased revenues from corporate income taxes, lower generalized consumption subsidies and higher taxes on property and the environment. Tax changes along these lines, combined with the universal nature of the social insurance programs that they would finance, would result in more redistribution towards low-income households and more jobs in higher-productivity firms.

Put differently, from a taxation perspective, the issue is not whether firms should be on or off the hook for the social insurance of their workers. The issue is that firms should be on the hook for those components of social insurance where they play an inherent role; that in parallel higher-income individuals should be on a more effective hook than at present; and that taxes should also contribute to increase efficiency in addition to raising revenues.

4.12 Income Inequality

This section considers the joint impact of our social protection and taxation proposals on income inequality. Again, we do not focus on any country. Rather, we make remarks that are hopefully relevant to all, but that are inevitably qualitative; as before, much depends on parameter values.

4.12.1. Measuring the impact of tax and social protection policies on inequality

We begin the discussion by making reference to relation (5), for convenience repeated here as (17), where we highlight the elements in L(.) and T(.) that change between the *status quo* and our proposals (identified respectively with the superscripts 0 and 1). Note that we ignore minimum wages L_{MW} since so far nothing was said about them. We return to discuss them immediately below, but at this point it is best to ignore them. We also ignore any changes to the institutions charged with enforcing labor regulations L_{ENF} or taxes T_{ENF}. With that, we have:

(17a) [(I.H, K); $\mathbf{T}$; $\mathbf{E}^0$\{ L(L_{CSI}^0, L_{JS}^0,., L_{NCSI}^0), T(T_{PIT}^0, T_{CIT}^0, T_{VAT}^0, T_{OTH}^0, T_{TRN}^0), M(.) }]

simultaneously determine [$\mathbf{R}^0$, $\mathbf{TFP}^0$, $\mathbf{SP}^0$, $\mathbf{Y}^0$]

(17b) [(I.H, K); $\mathbf{T}$; $\mathbf{E}^1$\{ L(L_{CSI}^1, L_{JS}^1,., L_{NCSI}^1), T(T_{PIT}^1, T_{CIT}^1, T_{VAT}^1, T_{OTH}^1, T_{TRN}^1), M(.) }]

simultaneously determine [$\mathbf{R}^1$, $\mathbf{TFP}^1$, $\mathbf{SP}^1$, $\mathbf{Y}^1$]

In words, even though the economy has the same factor endowments and technology, and the same "market conditions" M(.), changes in the realms of "entrepreneur-workers relations" L(.) and "taxes and transfers" T(.) impact resource allocation **R**, aggregate productivity **TFP**, access to social protection **SP** and market incomes **Y**.

Clearly, impacts would be substantial because changes in L(.) and T(.) would be measured in various points of GDP. Previously we argued that **TFP** would increase and that social protection **SP** would be more inclusive. In turn, higher **TFP** implies that $\sum \mathbf{Y}_i^1 > \sum \mathbf{Y}_i^0$. Put differently, GDP would increase and, as a result, so would average per capita income.

What would be the impact of the proposed changes on inequality? From (17a) and (17b) we can compute the Gini coefficient of market incomes, $G(\mathbf{Y}^0)$ and $G(\mathbf{Y}^1)$. Now, even though average per capita income would be higher, it is extremely difficult to tell what would happen to inequality. Lowering L_{CSI} and L_{JS} would increase formal employment and wages, particularly among low-income workers; this would tend to lower inequality. But profits in many micro and small informal firms would fall or even disappear, while profits in formal firms would increase. Further, the wage distribution would widen because as more productive formal firms expanded, the demand for workers with more years of education would strengthen—an effect that would tend to increase inequality. Potentially, the net effect could be $G(\mathbf{Y}^1) > G(\mathbf{Y}^0)$, and, in some cases this outcome may be welcomed, particularly if the driver is an increase in the returns to education.

Inequality of market incomes, however, is not as central to social welfare as inequality of disposable incomes, $G(\mathbf{Y} - \mathbf{t})$. This is particularly so under our proposals, which place a much greater emphasis on larger universal and targeted transfers financed by higher taxes. In this context, let relations (18a) and (18b) summarize the components of our proposals that can be captured in **t**: contributions to social insurance L_{CSI}, non-contributory programs L_{NCSI}, personal income taxes T_{PIT}, consumptions taxes T_{VAT}, property taxes and generalized subsidies on energy and the like T_{OTH}, and universal and targeted transfers T_{TRN}.

$$(18a)\ \mathbf{t}^0 = (\ L_{CSI}{}^0,\ldots,\ L_{NCSI}{}^0;\ T_{PIT}{}^0,\ldots,\ T_{VAT}{}^0,\ T_{OTH}{}^0,\ T_{TRN}{}^0\ )$$

$$(18b)\ \mathbf{t}^1 = (\ L_{CSI}{}^1,\ldots,\ L_{NCSI}{}^1;\ T_{PIT}{}^1,\ldots,\ T_{VAT}{}^1,\ T_{OTH}{}^1,\ T_{TRN}{}^1\ )$$

Contrasting $\mathbf{t}^1$ with $\mathbf{t}^0$, wage-based contributions for social insurance would be lower and there would be no social insurance programs for workers conditional on informal employment, so that $L_{CSI}{}^1 < L_{CSI}{}^0$ and $L_{NCSI}{}^1 = 0$. Very importantly, $L_{CSI}{}^1$ would exclude contributions for retirement pensions, eliminating the implicit tax when workers do not qualify for one because they fail to comply with the requirements and the implicit subsidy for those that do comply when contribution rates are below the actuarial value of benefits. Although these implicit taxes and subsidies are very difficult to measure, conceptually the point here is that they would cease. Equally, $L_{CSI}{}^1$ would also exclude contributions for health. Because $L_{CSI}{}^1$ would only include firm contributions that provide direct benefits to workers with no cross-subsidies, it would be neutral from the point of view of redistribution.

Personal income taxes $T_{PIT}{}^1$ would be very different from $T_{PIT}{}^0$. First, they would be higher. Second, they would include earmarked contributions for life and disability insurance and retirement pensions. Third, they would include subsidies for poor workers also earmarked for the same purposes. Fourth, they would also encompass many middle- or even high-income

individuals that currently take refuge in special regimes.[82] All in all, T_{PIT}^{1} would be substantially more progressive than T_{PIT}^{0}.

In some countries, consumption taxes T_{VAT} would be higher (or have fewer exemptions). The same could be true for other taxes T_{OTH} (like on property), while generalized subsidies for electricity or gasoline could be lower. Given the generality of the discussion, the distributional impact of these changes is very difficult to ascertain. Eliminating exemptions to the VAT or generalized subsidies would extract more resources from higher-income individuals, which is desired, but would also reduce disposable income of low-income ones, which is not.

In parallel, T_{TRN}^{1} would be larger than T_{TRN}^{0} because universal transfers for health and non-contributory pensions would be substantially higher. Very importantly, these transfers would represent a much larger share of the income of low-income than of high-income households, so they would be very progressive. Ideally, the value of these transfers would exceed any additional taxes on low-income households as a result of any changes to the VAT or to generalized subsidies. Considering these universal transfers, and targeted ones for the poor through CCT-type programs, the progressivity of transfers would increase significantly.

We now compare $G(\mathbf{Y}^1 - \mathbf{t}^1)$ with $G(\mathbf{Y}^0 - \mathbf{t}^0)$.[83] In general, the presumption is that inequality of disposable income would be lower under our proposals because transfers are more progressive, and because more tax revenues are extracted from higher- income than lower-income households. In other words, inequality of disposable income would fall because $\mathbf{t}^1$ would be very different from $\mathbf{t}^0$ and much more redistributive. That said, it is relevant to consider the opposite case. This could result if inequality of market incomes increases and is not compensated by the greater progressivity of **t**.

Although unlikely, from the point of view of income inequality $G(\mathbf{Y}^1 - \mathbf{t}^1) > G(\mathbf{Y}^0 - \mathbf{t}^0)$ would clearly be an undesirable outcome, although not necessarily so from a broader social welfare perspective. The reason is this: relative to the *status quo*, low-income households generally—and poor ones especially—would have full access to social insurance and potentially higher disposable incomes, even though relative to higher-income households they do not gain as much. Although inequality of disposable incomes would be higher, in absolute terms low-income households would be better off. Further, inequality along other critical dimensions of household welfare—access to health care, income during old age, unemployment and other forms of insurance and access to better jobs—would certainly be lower.

We make three points. The first is that Gini coefficients (or other measures of income inequality) should not be evaluated on their own. It is not the same to have a high Gini coefficient in a low-productivity, low-growth economy that offers few opportunities for advancement and provides erratic and incomplete coverage against risks, versus having it in a faster-growing economy where everybody is well insured against risks and workers have more opportunities for advancement. Context matters.

---

[82] Ideally, special regimes in the corporate income tax would also be substantially reduced, increasing revenues from this tax. This is not captured in **t** in relations (17) because it cannot be imputed to individuals or households.

[83] Note that, although tempting, we cannot compare $G(\mathbf{Y}^0 - \mathbf{t}^0)$ with $G(\mathbf{Y}^0 - \mathbf{t}^1)$ because, as discussed in section 3.6, if **t** changes, **Y** also changes. Some could argue that the comparison is a first-order approximation. But this line of reasoning would be flawed because changing L(.) and T(.) as proposed would have large impacts on occupational choices, the earnings distribution, firm profits and returns to education, making $\mathbf{Y}^0$ very different from $\mathbf{Y}^1$. In fact, the comparison could be quite misleading.

The second is that from the angle of inequality of disposable incomes, we need to focus on the impact of all taxes and the social transfers they finance—the complete tax-cum-social-protection package, so to speak (which is only partially approximated by relation (18)). Analyses of the progressivity or regressivity of individual taxes, while informative, are insufficient to assess their desirability. At times, measures to increase revenues like increasing the VAT or reducing generalized subsidies are ruled out because of their negative impact on low-income households. But this judgement should be contingent on what the revenues would be used for. If the revenues are used to improve social protection though expanded coverage of social insurance and more progressive transfers, and if some of the factors that obstruct the functioning labor markets are removed, low-income households could be better off. Again, context matters.

The third point is more general. From a normative perspective, it is difficult to establish a socially optimal value for G(**Y** – **t**). Given the pervasive inequality that characterizes the region, it is often assumed in policy discussions that the lower G(**Y** – **t**), the better. But clearly this cannot be true always. Unlike poverty, where many would argue that P(**Y**, $\mathbf{Y}^{cutoff}$) = 0 is a desirable outcome, not many would argue that G(**Y** – **t**) = 0 is desirable. Further, income inequality is driven by many more policies beyond social protection (education, competition and so on). In this context, three questions are relevant. How much of the effort to reduce G(**Y** – **t**) should fall on each? Is there a socially desirable value of G(**Y** – **t**)? What should policymakers aim for? We do not have answers to these questions, but think it is important to raise them to place the discussion of the relation between inequality and social protection in a broader context.

4.12.2. Minimum wages and income inequality

The discussion has so far ignored minimum wages. What role should they play in the tax-cum-social protection configuration proposed so far? To discuss this question, it is useful to return to Graph 11. On one hand, our proposals would make the labor market more competitive, reducing search costs and other frictions that create rents and reducing the space for the minimum wage to increase wages with small impacts on unemployment or informal employment. On the other hand, they would increase formal employment, particularly among low-wage workers, and thus amplify the set of workers who could potentially benefit from the minimum wage. It is hard to ascertain the balance, particularly because the answer depends critically on where the minimum wage is set relative to the wage distribution.

Because the labor market will never be perfectly competitive, it is clear that the minimum wage should not be set at zero. But, that said, at what percentile of the wage distribution should it be set? It is extremely difficult to provide an answer because there is a simultaneity issue that needs to be recognized:

- on the one hand, the wage distribution that results from our proposals could be very different from the observed one;
- on the other, the wage distribution that results from our proposals depends on the minimum wage.

In fact, to answer the question, one would need to first observe the wage distribution that results from our proposals without the interference of the minimum wage and then judge, based on the degree of frictions remaining in the market, what its optimal level should be. This, of course, is not possible.

In our view, however, what is critical is not so much where the minimum wage is set, but the fact that it does not obstruct the increased formality and higher **TFP** being sought by our proposals, particularly poor workers' access to formal jobs. This issue merits careful attention because we consider that these objectives should take priority over an attempt to lower G(**Y**) through a high minimum wage and because our proposals would in all likelihood lower G(**Y** – **t**) without recourse to the minimum wage. As a result, it may be best to set a minimum wage in the low range of the wage distribution, at least until tax and social protection policies are reformed along the lines suggested above and formal employment represents a much larger share of the labor force.

Put differently, we argue for a mix of caution, context and pragmatism. The discussion of minimum wages needs to be carried out considering all tax and social protection policies, not as a self-contained policy that can be set independently of the rest. Our proposals for UHI would give low-wage workers better health services than they have today and thus indirectly more disposable income. Our proposals for EUI and for negative income taxes earmarked for life and disability insurance and savings for retirement, would strengthen their bargaining position in the labor market without recurring to a high minimum wage.

In other words, there are alternative mechanisms to strengthen the bargaining position of workers *vis-à-vis* firms that can be equally effective but less troublesome from the point of view of efficiency. In fact, for most poor workers, finding longer-term formal jobs with more productive firms may over their lifetime be substantially more valuable than an erratic sequence of short-lived job, even if some of those jobs pay higher wages because of the lighthouse effect associated with a high minimum wage.

As discussed in the previous chapter, minimum wage discussions in the region are complex because at times two objectives are not clearly separated:

- to redistribute rents from firms to workers in labor markets where frictions create rents; and,
- to provide low-wage workers with a basic living standard.

In the context of Latin America, we consider that the second objective should be carried out mostly through the universal nature of social insurance, targeted transfers and negative income taxes, with minimum wages taking a secondary role for at least some time.

Paradoxically, the first objective could be more effectively pursued once there is less self-employment and formal firms are larger and employ a bigger share of the workforce. In those circumstances, the sharp trade-offs currently created by minimum wages would lessen and be similar to those faced by countries in the OECD. At that point, a suitably chosen w^{min} could be a useful complement in the fight against inequality. But most countries in Latin America are not there yet.

On the other hand, recall from Graph 11 that the minimum wage may also serve as a reference point for other policies, like indexing of pensions, floors for pensions or contributions for L_{CSI} for the self-employed or criteria to qualify for targeted poverty programs. These uses of the minimum wage create unnecessary difficulties and are not inherent to its core function.

We thus argue that, independently of its level, minimum wages policies need to be simplified. In general, they should not vary by the size of the firm or sector of the economy. It would be better to set them on an hourly rather than a daily basis and allow hours of work to vary. It would also be important to consider the different contexts of urban and rural areas, and

perhaps make special provisions for young workers to account for their lack of experience and costs of training (see Doruk et al., 2019). Finally, and very importantly, it would help a lot if they focused only on active workers and were not used as thresholds for other policies, avoiding using one instrument for two objectives.

4.13 Universal Social Protection and Strength of the Rule of Law

Here we discuss some implications of our proposals for the strength of the rule of law. Chapter three pointed out that in most countries, entitlements to social insurance are currently limited to dependent workers. With a few exceptions, like Argentina and Brazil, segmentation is inherent to the legal design; and even there, non-compliance and special regimes mean that not all workers enjoy equal entitlements to social protection.

4.13.1. Better laws

Our proposals are guided by a fundamental principle: all the relevant population should be entitled to protection from the corresponding risks and all should be required to contribute. Legally mandated segmentations would end. For the purposes of social protection, everyone would be equal under the law.

Laws underpinning social protection would be simpler and fairer. The plethora of complex and at times contradictory requirements to access this health program, that pension program or whatever program would disappear. Contributory pensions would be proportional to individuals' savings efforts. Nobody who contributed would be denied a pension or access to health services after retirement because time or amounts contributed were below some arbitrary thresholds. Individuals with similar earnings and working trajectories would have similar pensions. Simpler laws would be easier to administer, limiting the discretionary power of those enforcing them.

Firms' obligations would extend to all the workers they engage with, not only those with relations of dependency and subordination. Because social insurance laws would be more encompassing, the space for firms to elude them by arbitraging between contracts or subcontracting some tasks would be reduced. In addition, the incentives to evade would be reduced because contributions would be smaller and more valued by workers. In parallel, requiring all self-employed workers to contribute would also narrow the space to elude or evade.

Our proposals for protecting workers when they separate from firms would also reduce the space and incentives for litigation currently associated with the distinction between just and unjust causes of dismissal. Unjust causes would be circumscribed to firms' misbehavior like gender or ethnic discrimination. Changes in market conditions or technical innovations would be just causes. In parallel, workers would rarely have to sue firms to obtain their compensations because the bulk of them would be pre-paid before they separate from their jobs. The uncertainty, reputational and legal costs and sometimes corruption confronted when they exercise their labor rights would diminish. For the majority of them, what in Latin America is sometimes called labor justice would cease to be the largely unfilled promise of a long, complex and at times unenforceable labor code.

Eventually, as citizens experience a narrower gap between the *de jure* letter of social protection and labor laws and their *de facto* meaning, some tacitly accepted social norms—in particular the tolerance and at times celebration of evasion of social insurance contributions

and labor regulations—would give way to a recognition that it behooves everybody to comply with the law. Because disregard for the law is deeply engrained in some countries in the region, this would be a gradual process. But the important point is that our proposals would widen the doors for this process to take place.

Tax reforms required for universal social insurance also offer opportunities to strengthen the rule of law. Loopholes and special regimes in income tax laws would be substantially reduced or ideally done away with. Eliminating or narrowing exemptions in the VAT would reduce opportunities for arbitrage. A widening tax base would make it increasingly difficult to evade. These measures would enhance the perception that everybody is paying their due, gradually increasing tax morale and, more generally, the perception that everybody is treated equally.

4.13.2. Better enforcement of the laws

Although better laws would be easier to enforce, the region still needs to improve the functioning of the institutions in charge of collecting taxes and social insurance contributions and enforcing labor regulations. The combination of simpler laws, lower incentives to evade and greater penalties for doing so would bring the region closer to the situation observed in most countries in the OECD: laws are by and large complied with. This will not happen quickly, but the pace of change can be accelerated if, in parallel to reforming social protection laws, countries invest more resources in strengthening labor ministries, agencies in charge of collections and labor tribunals. These investments are rarely considered part of social protection spending, but they should be.

4.13.3 The end of informality?

Finally, our proposals would set the basis for the end of informality. Indeed, under our proposals the words formal and informal would lose operational meaning. There would be some obligations for firms, some obligations for income earners and some obligations for taxpayers. Critically, all obligations would be universal. In this context, and in contrast to the *status quo*, there is no need to use the words formal and informal as distinct from the words legal and illegal. Whoever is not complying—firm, income earner, taxpayer—would be engaging in illegal behavior, period. Euphemisms like “informal salaried employment” would eventually cease to be part of the language of social protection, to be renamed as it should, “salaried workers hired by firms violating the law.” Unlike the *status quo*, such violations would only have implications for the protections associated with firms, but not for health or retirement pensions.

Full compliance by all is not observed anywhere in the world, and even with our proposals, Latin America would be no different. But gradually, Latin America could approach the situation of other regions of the world where compliance is larger and where evading the law is shunned. Living at the edge of the law or on the wrong side of it would no longer be part of “the way things are” (because “that is the way they have always been”). Participation in social insurance—contributing to and benefiting from—would be expected from all members of society, an expression of citizenship.

4.14 Towards More Encompassing Social Protection

The discussion in this chapter has focused only on insurance and redistribution. But social protection systems should be more encompassing, including policies that were placed in the border between social protection and investments in human capital in Graph 1—in particular policies to improve labor training and matches between firms and workers (i.e., active labor market policies) and policies to facilitate women's participation in the labor force (like childcare). They both directly impact labor markets, workers' incomes and **TFP**, even if they can also be classified as investments in human capital.

Thus, our proposals should be seen as the core elements of a more encompassing social protection system, which needs to be complemented with other policies that are not discussed here. These policies should also be guided by the principle of universality with respect to the relevant population, even if coverage is built up gradually for budgetary reasons.

That said, social protection should not be an ever-expanding set of programs without an underlying logic. Graph 17 hopefully captures a key idea: a well-functioning social protection system is not a collection of programs working in isolation. It is a set of policies and programs that jointly provide sufficient protection against risks and is effective in combating poverty and reducing inequality while contributing to growth and the strengthening of the rule of law.

Graph 17: Constructing More Encompassing Social Protection Systems

Source: authors' elaboration.

This set need not be the same across the region, of course. Some countries may make greater efforts to redistribute through health programs relative to pensions; within pensions, some may rely more on higher contribution rates for contributory ones and some for lower rates compensated by more generous non-contributory pensions. In others, unemployment insurance can be less generous because compensations at times of separation are higher and so on. There clearly is some substitution across policies. But what matters is that all these programs, together with the taxes deployed to finance them, form a coherent whole. In other words, what matters is that, in each country, the pieces of the puzzle fit together, even if pieces are different sizes across countries. Furthermore, in each country additional pieces

need to be incorporated in response to emerging needs.[84]

Social protection also needs to be more encompassing of different lifestyles. The working age population is traditionally divided between those that are active and those that are out of the labor force. But as with the formal and informal labels, these are not permanent characteristics of individuals, but transitory status during their lifetimes. As cultures evolve, traditional roles as to who does what when are changing. Social protection should support that change, and our proposals do so. Regardless of how often individuals transit in and out of the labor force, they would always have access to health care and a non-contributory pension and never lose their savings for a pension carried out while they are in the labor force.

Social protection must also be more encompassing regardless of how technology evolves. In this essay $\mathbf{T}$ has been a black box in the background from which firms chose specific technologies with a given TFP_i. But changes in $\mathbf{T}$ are weakening the division between dependent and non-dependent workers and erasing the sharp line between those that are paid salaries and wages and those that are paid on a piece-rate basis or through commissions or rent-sharing arrangements. It is also blurring the line between working at home and doing so in the factory or the office. Many of these trends have been present for some time but are accelerating as a result of the so-called gig economy. In fact, based on current evidence, one can reasonably expect that economic activity based on firm-worker interactions mediated by digital platforms will increase in the future, blurring the lines between self and dependent-employment and the distinctions between salaries, wages and other forms of remuneration. Individual workers will engage with many firms at the same time, in turn blurring the employer-subordinated worker relation.

The ($L_{CSI} + L_{JS} + L_{MW}$) regime is ill-suited to accommodate these trends, developed as it was over a century ago when the expectation was that the vast of majority of workers would have a single job with a single firm, working in a fixed location where efforts and hours of work could be monitored and where they would be paid a fixed remuneration per unit of time worked regardless of the output produced. This expectation was never really fulfilled in Latin America and, for the reasons just mentioned, will be even less so in the future. Social protection needs to ensure that individuals are always covered regardless of how technology evolves. If it does that, as we consider that our proposals do, then it will facilitate the adoption of new technologies, which will in turn open better opportunities for the region's increasingly educated labor force.

Social protection also needs to be more encompassing through time, that is, more permanent. Households need certainty that, regardless of the vagaries of their individual circumstances or the economic cycle, they will always be protected from key risks. Social protection should be a source of tranquility rather than, as is at present for many workers, a source of stress when they lose their job, they face labor litigation, the family business goes bankrupt, or the illness of a household member increases financial pressures on everyone. There will always be negative idiosyncratic or systemic shocks, but they should not always translate into setbacks, suffering and distress.

[84] An emerging need that is already here as a result of the region's demographic transition is associated with care for the elderly, beyond health and pensions; see Cafagna et al. (2019).

More encompassing social protection is probably insufficient to generate more social inclusion and a sense of basic fairness across countries in the region and to accelerate growth and strengthen the rule of law. But it is the best place to start.

5. Universal Social Protection is Possible in Latin America

5.1. Looking Back

Over 75 years ago, countries in the region began constructing their social protection systems. In all of them, the cornerstone was a combination of three elements: contributory social insurance programs, regulations on job stability and minimum wages—what we have called the "($L_{CSI} + L_{JS} + L_{MW}$) regime". There are differences across countries in aspects of this regime of course, but the similarities dominate. Earmarked wage-based taxes—sometimes called social insurance contributions, sometimes payroll taxes—would finance a combination of health, pension and other programs to protect workers against risks; regulations on stability would assure workers' permanent access to these programs; and minimum wages would provide a floor on their incomes. Urbanization, industrialization and overall growth would increase the share of the labor force covered by this regime, gradually expanding social protection to all. Clearly, other programs attuned to countries' circumstances complemented this regime but, looking back, what stands out is its centrality.

We now know that this regime did not deliver social protection for all—in fact, in most countries, not even for the majority. For some time, it was thought that macroeconomic mismanagement was to blame. But we have learned that this is not so. With exceptions, macromanagement in the region improved considerably over the last 25 years, but coverage of the ($L_{CSI} + L_{JS} + L_{MW}$) regime is still substantially below what is needed for effective social protection. It has also been thought that, beyond macromanagement, the problem is lack of growth. "If somehow growth accelerated, coverage of the ($L_{CSI} + L_{JS} + L_{MW}$) regime would expand" is something we hear often. But we now also know that there is a two-way relationship between social protection and growth, and that ignoring spurts from investment booms and positive terms of trade shocks, deficiencies in the former—prominently in the regime—partly explain why the latter is slow.

The regions' policymakers have not stood still in the face of the regime's incomplete coverage. Once the debt-provoked convulsions of the 1980s came to an end, starting in the early 1990s well-intended reforms were made everywhere to expand the coverage of social protection. And we should all welcome this, given the tremendous social costs imposed by those convulsions and the large inequities that pre-dated them.

But, by and large, policymakers did not expand coverage by tackling the deficiencies of the ($L_{CSI} + L_{JS} + L_{MW}$) regime; they did it by bypassing them. This is not to say that individual programs in the regime were not reformed. Some were, particularly in pensions, modifying the parameters of PAYG systems or, in some cases, replacing them with defined contribution

ones. Nor is it to say that some tinkering did not occur here and there with health and other programs in the regime. But it is to say that its underlying design was left basically intact. In all countries, wage-based contributions still finance a combination of programs to protect workers against risks, complemented by provisions on job stability and minimum wages.

Instead, policymakers expanded coverage through new parallel programs, sometimes under the label of non-contributory social insurance and other times under the label of targeted poverty programs. These programs proliferated everywhere and so did social spending, although not always with a solid tax base behind it. These programs were accompanied by measures promoting micro and small firms and sometimes subsidies to consumption, in the expectation that these policies would jointly expand the coverage of insurance, reduce poverty and mitigate income inequality.

Today, social welfare would be worse without these programs. Indeed, the problem is not with them on their own; in fact, some have been carefully evaluated and, in some cases, results are encouraging. The problem is that the plethora of non-contributory social insurance and targeted poverty programs, jointly co-existing with the (L_{CSI} + L_{JS} + L_{MW}) regime and special regimes for small firms, result in complex social protection architecture in terms of sources of financing and rules and conditions of access.

This essay has argued that the region is not well served by this architecture. There are untenable differences in terms of workers' entitlements to social benefits; similar workers making similar efforts are not treated equally by the law. Protection against risks is effective for only a subset of workers, with the rest—more often than not the majority—poorly and erratically protected at best. Redistribution is weak and sometimes in the opposite direction of what is desired, rhetorical displays notwithstanding. And, if this was not enough, key factors that determine productivity and long-term growth are negatively affected. Socially inclusive growth cannot result from this situation.

This is not to say that growth cannot occur unless social protection policy is reformed; it may. But as has happened in the region in the past, many will be left behind. And if policymakers increase spending in social protection while preserving its current architecture to avoid that outcome, they will slow down growth. Put differently, the region's social protection architecture is an obstacle to socially inclusive growth; if this type of growth is the *desideratum*, then reforming social protection needs to be at the top of the growth agenda.

Yes, the region needs to invest more in infrastructure and yes, it needs to expand access to internet and yes, it needs to ..... the list can be long. But, looking back, we have all learned that not everything can be done at the same time. Ours is not an argument against more infrastructure, or expanding the internet, or whatever is the latest "key" to growth. It is an argument about priorities.

The crisis associated with Covid-19 is very different from the debt crisis of the 1980s. The first is a truly exogenous event caused by a virus; the second was in large measure endogenous to the region, caused by poor macroeconomic management. But there is one dimension in which the first may be very similar to the second: it will be followed by important changes to social protection and taxation as countries struggle to contain the social damage, restore fiscal balances and resume growth. In this context, the early 1990s response to the lost decade that preceded it offers Latin America an extremely valuable lesson. In the context of the current social protection architecture, raising taxes to fund more social spending will not result in a shared and sustained prosperity. If we can learn from the past, Covid-19 may mark a break with it.

5.2. Looking Forward

There is a better approach. The principle of universality with respect to the relevant population can help the region escape from its present predicament, as long as universality is simultaneously pursued in three complementary dimensions: coverage, quality of service and source of financing. This principle has different implications for programs in health, retirement, death and disability pensions, protections against accidents at work and loss of employment, and poverty; but in all cases it induces better social outcomes than present ones. A social protection architecture built around this principle offers the region a route to increase social spending while strengthening the foundations of long-term growth, a more promising option to combat poverty and a path to enhance social inclusion and strengthen citizens' regard for the rule of law.

Universality with respect to the relevant population provides a guide as to which policies and programs should be pursued, including proposals for new ones like a universal basic income. It also provides a guide as to which policies should be financed from general tax revenues, which from contributions by workers and which from contributions by firms. If social protection is going to be more inclusive, fiscally sustainable and growth-enhancing, it needs to be more than a set of outdated laws and an ever-expanding collection of programs, financed from whatever combination of taxes and contributions can be cobbled together. It needs to ensure that everybody is protected and that the programs that are put in place provide incentives to firms and workers to increase productivity.

Universality opens new avenues to combat poverty. For the first time, the poor would have access to insurance under the same conditions as everybody else, including services of equal quality. Equally important, the obstacles stemming from social protection policies that currently block their access to more productive jobs would be removed. Because the architecture of social protection would be more favorable to increased productivity and faster growth, targeted income transfers, aside from playing a secondary role, would more likely be transitory.

Universality would represent a major change in social entitlements. For the purposes of social protection, the norm would be equality under the law. Social protection laws and institutions would contribute to phasing out the formal-informal segmentation of the economy, strengthening social inclusion and, gradually, increasing citizen's regard for the rule of law.

Of course, the principle of universality with respect to the relevant population is not a rigid blueprint. Within the proposals that we advocate, there are variants and combinations of policies, as well as parameter values for each, whose pros and cons depend on the characteristics of individual countries, including political preferences for taxation and social solidarity.

The dysfunctional social attitudes towards the rule of law in parts of the region have complex causes. One of them may be perceptions of inequality of opportunity or absence of a levelled playing field. But these perceptions are not cast in stone. The principle of universality with respect to the relevant population may help modify them to the extent that it improves three critical aspects of households' welfare, particularly low- and middle-income ones: access to better jobs, access to more effective insurance and access to better quality services. This is probably insufficient to reverse deep-rooted attitudes, but it could be the beginning of a virtuous circle in which wider social inclusion generates a greater sense of belonging and more social cohesion overall. It is difficult to think of societies with strong rule of law and high social capital with fragmented and ineffective social protection systems.

Socially inclusive growth has eluded Latin America partly because social protection policies undermine some of the long-term determinants of growth, particularly productivity. There are other reasons of course. An important one in some cases has been the recurrent inability to manage international commodity cycles. When conditions are transitorily favorable, governments overspend, only to face costly reckonings later on. In part, these policy responses may be driven by understandably large pent-up social demands that are extremely difficult to postpone. When the going is good, it is hard to ask people to save for a rainy day, when for many every day is rainy. However, if low- and middle-income households have access to effective social protection and if they experience steady progress in their material well-being, they may have longer horizons and larger stakes in prudent macromanagement. Unsustainable expansionary policies have many roots, but inequities, exclusions and deprivations are among them.

Indeed, one could argue that the cycles between what in the political lingo of the region are called neoliberal and populist regimes at least in part reflect an underlying distributional conflict that has not been fully resolved. Our proposals are no guarantee against these cycles, of course, but they may make them less likely because they open a route to growth that is both socially inclusive and macroeconomically sustainable.

5.3. A Complicated but Not Impossible Political Economy

Restructuring social protection along the lines suggested in this essay is an ambitious task. Indeed, some may think that this task is impossible and characterize our proposals as little more than an academic exercise. Realists may argue that the existing social protection architecture should be seen as the equilibrium outcome of many countervailing interests, leaving little if any room for blueprints for reform. Countries rarely carry out changes like the ones proposed here and attempts to do so should at best be characterized as quixotic.

We disagree. No doubt, interests are central determinants of countries' social protection architecture. But they are not the only ones; ideas also matter. In our view, these architectures are shaped by the overlapping forces of interests and ideas, although inertia and path dependence also play a role. Clearly, some groups profit from the *status quo*. But others may be opposed to change it even if they do not profit from it because of their views of the world; for them the issue is ideas, not interests.

We do not underestimate interests; restructuring social protection faces a complex political economy everywhere, although obstacles differ across countries. And while a careful analysis exceeds the scope of this essay (and our abilities!), some general remarks may be useful.

5.3.1. Taxation

A mechanical exercise like the one presented in chapter four, measuring the fiscal costs of restructuring social protection and identifying sources of revenue, can establish that, *mutatis mutandis* across the region, there is room to raise additional revenues to meet costs. This is critical if social protection is to have, as it should, a stable revenue base, beyond the vagaries of international commodity prices.

But evidently the issue is far more complex than that. Behind every tax loophole, special regime and exemption there is a group that benefits. These groups, some very politically influential, would in all likelihood oppose more taxes; there is nothing new here. After Covid-19 some among them may understand the need to strengthen social protection. Moreover, as we briefly touch on below, coalitions can be built to weaken the opposition of others, particularly if society as a whole perceives that more taxes will not be used to finance more of the same, but rather an alternative that can simultaneously result in more social cohesion and faster growth.

However, in Latin America many who would in principle benefit from higher taxes may also oppose them. Perceptions of corruption, lack of transparency and weak accountability translate into distrust of government and low willingness to pay taxes.[85] This is compounded when the quality of public services is deficient. These are powerful forces conspiring against more taxation, and interest groups that would be hurt by higher taxes often take advantage of them to sabotage tax reforms.

But these forces are not immutable. Fighting corruption and improving the quality of services is urgently needed in the region, with or without a change in social protection. But a change in social protection may provide an additional impetus to do so; in fact, it may be the driving force of an effort to renew the institutions delivering services and ensure the probity of government management. It is hard to think that taxes could increase by various percentage point of GDP if these measures are not taken, even if governments claim that the extra revenues will fund a universal system of social protection. Put differently, restructuring social protection can open a path to escape the vicious circle of low trust, low quality of services and low willingness to pay taxes.

Opposition to more taxes may also arise from perceptions that social spending keeps a segment of the population permanently living on public welfare. However, these perceptions are not immutable either. Societal preferences for redistribution may be changed if those paying taxes consider that social protection policies are more effective than current ones. In this context, the novelty in our proposals is that more taxes would not be channeled to fund more of the same policies; they would be channeled to transform them. More taxes would fund a decisive change towards universality in access to social protection and increased productivity—a change that would reward better the efforts of poor and low-income households to improve their situation. In parallel, more taxes would fund stronger incentives to comply with the law, combat evasion and widen the tax base.

Thinking about tax reforms needs to change. At times it is thought that taxes can be raised as long as low- income groups, or those living in poverty, are compensated. But this thinking fails to notice that the higher targeted income transfers, the higher the obstacles for the poor to escape poverty. At other times it is thought that taxes can be increased with an expansion of social programs for those who are informally employed. But this thinking fails to notice that these measures will increase the incentives to be informal and erode the tax base, increasing the tax burden on those already paying. It is time for the region to move away from tax reforms with compensations that only seek to increase the tax burden while leaving almost

[85] Using data from opinion surveys carried out by Latinobarómetro, the OECD finds that in Latin America, tax morale is low and has fallen in the last few years. In 2015 only 48 percent of households considered that it is never justified to evade taxes, and 27 percent considered that it is justified; OECD et al. (2020).

everything else intact to more systemic reforms that also consider their impact on firms' and workers' incentives, in addition to their ability to raise revenues.

5.3.2. Special interest groups

Opposition to change may also occur for reasons unconnected to taxes. At present, in some countries a few powerful and vocal groups enjoy special pensions or privileged access to health services, even within contributory systems—teachers, the police, public sector workers or members of elite institutions like development banks. These situations are partly an inheritance of the regions' corporativist politics and partly a result of governance failures, including insufficient transparency and accountability.

Consider, for instance, unions. In Argentina, they are involved in managing contributions for health services and could oppose universal health insurance if it implies losing the benefits from this role. But ways could be found to protect them from losses. Unions are not an absolute impediment to universal health insurance, as Brazil attests. More generally, unions everywhere may be more willing to accept universal health insurance if they are assured that the quality of services received by their members will not decline and that in fact, unlike the situation today, they will receive the same care while they are unemployed or temporarily leave the labor force. Similarly, unions might be more willing to change cumbersome and often-violated job stability regulations if they are assured that a better option is available through the combination of cause free termination payments, enhanced unemployment insurance and dismissal payments proposed before. The key point is to convince them that their members would be better off. This may take time and lots of dialogue but is not impossible because they would be better off.[86]

In Mexico, among other countries, unionized public sector workers enjoy better pensions than private sector ones, even when the latter are unionized. These situations may have to be grandfathered, with changes applying only to new members. But even if this is not possible, reforming public sector pensions is not required to restructure social protection, at least not initially. The bulk of the labor force is in private firms, including one-person firms, and it is there where our proposals would have the greatest benefits, both in terms of social protection and productivity.

Potential opposition from unions needs to be put in perspective. Latin America has low union density compared to developed countries.[87] In part, this is a result of large informality; it is hard to organize the self-employed or those working in very small firms, many of which are not even registered as such. These facts highlight an observation that is sometimes lost: unions could be large beneficiaries of the restructuring of social protection that we propose. Minimum wages can enhance the bargaining power of workers versus firms, but their effects pale compared to the strength that unions could derive from a situation where workers are

[86] This is not to deny that there may be occasions where the interest of the union and its members are not fully aligned, but to say that if indeed members are better off, it is more difficult for unions to resist change.

[87] Ronconi (2021) uses data from the International Labour Organization to compare average union density—the ratio of unionized workers over the labor force—between Latin America, 20 percent, and developed countries, 30 percent. That said, there is large variance in the Region: 30.1 percent in Argentina, 18.1 percent in Brazil, 15.8 percent in Chile, 14.2 percent in Mexico, 9.4 percent in Colombia, 9.3 percent in the Dominican Republic and 4 percent in Peru (unfortunately, no data are available for Ecuador and Honduras).

more concentrated and therefore more easily organized because there are more and larger formal firms. More formality is probably the deepest structural change that could strengthen the bargaining power of unions and workers.

Aside from reinvigorating unions, our proposals could also change the dynamics of their interactions with firms. Surely, negotiations on wage and non-wage benefits would continue to be central, but they could expand to issues where firms and workers can both gain: training programs, coalitions to improve the quality of services, the fight against corruption. OECD countries have unions, and they also have much better social protection systems; they are not inherently incompatible.

Other special groups may also oppose change, although much depends on the details. Private administrators of pension funds may be one of them, and they can be powerful, particularly when they are part of large financial conglomerates. Our proposals are compatible with both PAYG or defined contributions systems through individual accounts. But in the latter case, the fees and commissions charged to manage these accounts are in some cases unjustifiably high, and they need to be reduced as part of the changes that we propose. Financial groups would oppose this, but their opposition could lessen if many more people than today were saving for their retirement in their individual account and for longer spells of time. Yes, margins would decline, but volume would compensate.

More generally, the universal nature of our proposals could change the political economy behind low-quality services and poor regulation, particularly in health and retirement pensions. In the case of health, everybody would contribute through taxation; contributions would not be diverted to separate systems as happens currently in Argentina, Chile and, to a lesser extent, Peru, nor would there be two separate systems, one for formal workers and one for informal as occurs in the Dominican Republic, Ecuador, Honduras and Mexico. Because nobody would be allowed, in Hirschman's (1970) framing, to vote with their feet and exit from contributing to the unified public health system, more groups, including higher-income ones, would have the incentives to exercise a strong voice to improve its quality, all the more so if expenditures for private health insurance are no longer deductible from personal income taxes.

Something similar would occur with retirement pensions. Under our proposals, the obligation to save for one would apply to all income earners, not only workers. The fact that high-income entrepreneurs and professionals like doctors and lawyers contribute to their pension through the same mechanism that all others do would help to focus their attention more sharply on the laws that regulate pensions and the functioning of the associated institutions—again, all the more so if contributions to private plans are also not tax deductible.

Still, changes to taxation would likely be opposed by middle- to high income households, for whom social insurance may not be high on their agenda. But changes to taxation could take place in a gradual process, taking advantage of other objectives that are high on their agenda, like improving the environment and combating climate change, in which case taxes on negative environmental externalities or fossil fuels could be part of the tax effort. The process of change, of course, would also require building trust, which would in turn require a credible fight against corruption. The key point is to move, even if gradually, from a low trust, low tax and low quality of services equilibrium to a higher trust, higher tax, higher quality of services one.

The list of groups that oppose change varies from country to country. But, in all, the list of those that would benefit from it is much, much longer. Precisely because of this, in our view the relevant question is not whether our proposals are realistic; the relevant question is how

to navigate the political economy in each case, complex as it may sometimes be, so that the required changes can be brought about.

What is the alternative, really?

5.4 The Centrality of Ideas

A distinction needs to be made between a vision of what a country's social protection architecture should look like and a political assessment as to how that vision can be achieved. The latter is indispensable of course. The transition from the *status quo* towards a social protection architecture along the lines proposed here would be complex in all countries, although different in each. Initial conditions differ as well as administrative capabilities; the same holds for political contexts. But everywhere, restructuring social protection would be a process that faced resistance from various groups and, almost inevitably, unexpected negative shocks along the way. All this can derail or deviate change.

For these reasons, a vision is *sine qua non*. In fact, one could argue that the absence of one is partly to blame for the current situation. Without a vision, social protection is little else than the inertia of long-dated policies and programs and a continuous addition of new ones, sometimes in reaction to immediate needs, sometimes in compliance to the whims of the president in charge, sometimes in imitation of policies in more advanced countries with very different realities, and sometimes in response to the latest fad in academic or international policy forums, with little coherence between them. In other words, the *status quo.*

That is why, regardless of the vagaries and speed of the process of change, it is important to have clarity as to what policymakers should aim for. Policymakers need a compass to navigate through what will sometimes be choppy waters. A vision of the desirable social protection architecture is, we believe, an indispensable compass to navigate successfully. Independently of whether individual reforms are carried out rapidly or slowly, what is critical is that they are compatible with each other and jointly advance the goal of universality sketched here.

Ideally, countries could first set the contours of their desired social protection architecture, along the lines of the discussion in chapter four, but of course with the adaptations and modalities pertinent to each. With that, countries could map out a sequence of inter-temporally consistent reforms that, overtime, get them closer to the desired endpoint. The direction of individual changes, given other changes already made or planned, is as important as their speed. Whether the transition takes 1, 5 or 10 years is less relevant than the fact that countries are consistently working in the right direction. Of course, it will never be as simple or as linear as that. Almost inevitably, there would be setbacks and mistakes, but a compass can serve to reset the direction of reform once conditions change.

Reforms can create winners, modify the balance of power and weaken those who support the *status quo* and open opportunities for further reforms. They can result from a broad political understanding among all concerned, from coalitions, elections or exogenous shocks that create opportunities. Reforms may begin with health programs, pensions or unemployment insurance and job stability regulations. This would likely vary from country to country, but the important point is this: when the political space for change appears, or is created, it is important to have clarity about what needs to be done. Ideas are central.

Ideas can change interests. At present, middle-income groups may oppose isolated reforms because they only perceive costs. They oppose eliminating generalized fuel subsidies because compensations are usually targeted at groups with lower incomes than theirs. They oppose tax increases because they perceive that they finance social protection programs that fail to address their needs. Even lower-income groups can oppose reforms that in principle would benefit them because they are weary of the negative effects of previous reforms.

Under these perceptions, reforms are very difficult. In some extreme cases, perceptions can be so negative that an individual reform results in civic unrest, which usually ends with whatever patch-up measure can be quickly assembled, leaving the desired reform waiting for a better day. In other cases, these perceptions only allow minor reforms, keeping the underlying architecture of social protection intact, going forward in one direction and backward in another one—all this while consuming policymaker's scarce political capital and creating reform fatigue as citizens see little if any gains. And yet in other cases, these perceptions end in policy paralysis. Whatever the outcome, under these perceptions, countries are stuck in an equilibrium where almost everybody loses.

An overall vision of the social protection system may change them. A proposal for a broad redesign might allow different groups to see what is in it for them and to understand that maybe they can be winners. Yes, individual components of the proposal might be undesirable to some groups, but the proposal as a whole may not; and what is undesirable to some might actually be desirable for others. All in all, a broad proposal for systemic reform might then be more acceptable and have more chances of success than isolated reforms, even if the broad proposal is implemented gradually. But for this to happen, members of society need to know where they are being led—what the endpoint looks like. They need to be presented with new ideas for them to consider that their interest might lie elsewhere.

A century ago, social protection systems were practically absent in Latin America. Countries began constructing them in reaction to various political demands. But the policies deployed were based on ideas which, at that time, could have easily been labelled as unrealistic: wage-based contributions for health and pensions coupled with minimum wages and job stability provisions for countries where over 75 percent of the population was rural, and with lower-income levels and weaker state capacities than today. We now know that this system did not deliver the expected results, but the point here is that it was based on ideas—on a vision of the world, if we may say so. And, at that time, that vision was bold.

The current architecture of social protection is a reflection of those ideas, surely modified along the way by interests and country specificities, but by and large still the same. These ideas have had their day. Improving the welfare of millions of people in the region requires that we all move on.

Interests will always be there. But it is useful here to remember Keynes. Policymakers may be more imprisoned by old ideas than by interests. New ideas can counteract the efforts of influential groups opposed to more taxes or to the phase out of historic privileges. But for that to happen it is first necessary to change the nature of the political debate. No to more taxes for more of what has not really worked; yes, to more taxes for an option that can achieve the twin objectives of universal social protection and faster growth. No to more isolated social protection reforms that deepen segmentations and perpetuate exclusions; yes, to a process of systemic reform aimed at including everyone. Reforms based on new ideas may still not happen, but for them to have a chance of happening, the political debate needs to change.

One can find reasons to be optimistic. The experience of the region shows that when a social need is identified, substantive reforms can happen if the proposals for reform are presented widely and openly to all. This was, for example, the case of Chile during the second term of President Bachelet, when social preferences were expressed for greater public spending in education and an important tax reform earmarked for these purposes was put in place. We do not judge the merits of those reforms, but we celebrate the fact that a Latin American society was capable of simultaneously realizing a significant tax and social policy reform. Experiences like this one are of great relevance. Political assessments of social reforms and attitudes towards taxation are not immutable; change can happen. Visions of what social protection policy should look like are relevant. Political economy is not everything—ideas matter.

It is time for a new course.

References

Acosta, P., Cruces, G., Galiani, S., and Gasparini, L. (2019). Educational Upgrading and Returns to Skills in Latin America: Evidence from a Supply-Demand Framework. *Latin American Economic Review*, *28*(1), 1-20.

Alaimo, V., Bosch, M., Kaplan, D., Pagés, C., and Ripani, L. (2015). *Jobs for Growth*. Inter-American Development Bank.

Alaimo, V., Bosch, M., Gualavisi, M., and Villa, J.M. (2017). *Measuring the Cost of Salaried Labor in Latin America and the Caribbean* (Technical Note IDB-TN-1291). Inter-American Development Bank.

Alfaro, L., and Kanczuk, F. (2020). América Latina: el problema de la productividad [Latin America: The Problem of Productivity]. *El desafío del desarrollo en América Latina: políticas para una región más productiva, integrada e inclusiva*. Corporación Andina de Fomento.

Alvarado, F., Meléndez, M., and Pantoja, M. (2021). *El sistema de protección social de Colombia* [The Social Protection System in Colombia] (UNDP LAC Working Paper No. 14). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Alvaredo, F., and Gasparini, L. (2014). Recent Trends in Inequality and Poverty in Developing Countries. In A. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution, vol.2*.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). The Elephant Curve of Global Inequality and Growth. *AEA Papers and Proceedings, 108*, 103-108.

Alzua, M.L., Cruces, G., and Ripani, L. (2012). Welfare Programs and Labor Supply in Developing Countries: Experimental Evidence from Latin America. *Journal of Population Economics, 26*(4), 1255-1284.

Alzua, M.L., and Pacheco, A. (2021). *Protección social, formalidad y subsidios cruzados: evidencia para Argentina* [Social Protection, Formality and Cross Subsidies: Evidence for Argentina] (UNDP LAC Working Paper No. 16). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Antón, A., Hernández, F., and Levy, S. (2012). *The End of Informality in Mexico? Fiscal Reform for Universal Social Insurance*. Inter-American Development Bank.

Araujo, M.C., Bosch, M., Maldonado, R., and Schady, N. (2017). *The Effect of Welfare Payments on Work in a Middle-Income Country* (Working Paper Series IDB-WP 830). Inter-American Development Bank.

Azuara, O., Azuero, R., Bosch, M., and Torres, J. (2019). *Special Tax Regimes in Latin America and the Caribbean: Compliance, Social Protection and Resource Misallocation.* (Working Paper Series IDB-WP 970). Inter-American Development Bank.

Azuero, R., Bosch, M., Cardoza, M., and Sánchez, D. (2019). *Productivity, Misallocation and Special Tax Regimes in the Dominican Republic.* (Working Paper Series, IDB-WP 1050). Inter-American Development Bank.

Bachas, P., Gadenne, L., and Jensen, A. (2020). *Informality, Consumption Taxes, and Redistribution* (Working Paper 27429). National Bureau of Economic Research.

Baez, J.E., and Camacho, A. (2011). *Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia.* (Policy Research Working Paper 5681). World Bank.

Bando, R., Galiani, S., and Gertler, P. (2020). The Effects of Noncontributory Pensions on Material and Subjective Well-Being. *Economic Development and Cultural Change, 68*(4), 1233-55.

Banerjee, A., Niehaus, P., and Suri, T. (2019). Universal Basic Income in the Developing World. *Annual Review of Economics, 11*, 959-83.

Banerjee, A., Hanna, R., Kreindler, G., and Olken, B. (2017). Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs. *World Bank Research Observer, 32*, 155-184.

Barinas, S., and Ñopo, H. (2021). *Políticas de protección social y laboral en República Dominicana* [Social and Labor Protection Policies in the Dominican Republic] (UNDP LAC Working Paper No. 18). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Benartzi, S., and Thaler, R. (2007). Heuristics and Biases in Retirement Savings Behavior. *Journal of Economic Perspectives, 21*(3), 81-104.

Benavides. J., Guzmán, C., and Perry, G. (2020). Infraestructura y desarrollo [Infrastructure and Development]. *El desafío del desarrollo en América Latina: políticas para una región más productiva, integrada e inclusiva*. Corporación Andina de Fomento.

Bernal, R., Eslava, M., Meléndez, M., and Pinzón, A. (2017). Switching from Payroll Taxes to Corporate Income Taxes: Firms Employment and Wages After the 2012 Colombian Tax Reform. *Economia*, *18*(1), 41-74.

Besley, T., and Kanbur, R. (1990). The Principles of Targeting. In V.N. Balasubramanyan and S. Lall (Eds.), *Current Issues in Developing Economics*. Macmillan.

Bérgolo, M., and Cruces, G. (2014). Work and Tax Evasion Incentive Effects of Social Insurance Programs: Evidence from an Employment-based Benefit Extension. *Journal of Public Economics*, *117*, 211-228.

Bérgolo, M., and Cruces, G. (2021). The Anatomy of Behavioral Responses to Social Assistance when Informal Employment is High. *Journal of Public Economics, 193.*

Berstein, S., Bosch, M., García Huitron, M., Oliver, M.L., and Altamirano, A. (2018). *Diseñando las pensiones del futuro en América Latina y el Caribe* [Designing the Pensions of the Future in Latin America and the Caribbean]. Inter-American Development Bank.

Blanchard, O., and Tirole, J. (2004). *The Optimal Design of Unemployment Insurance and Employment Protection (*Working Paper 10443). National Bureau of Economic Research.

Blank, R. (2002). Evaluating Welfare Reform in the United States. *Journal of Economic Literature*, *XL*, 1105-1166.

Bobba, M. Flabbi, L. and Levy, S. (2021). Labor Market Search, Informality and Schooling investments. Forthcoming International Economic Review.

Bosch, M., and Manacorda, M. (2010). Minimum Wages and Earnings Inequality in Urban Mexico. *American Economic Journal: Applied Economics*, *2*, 128-149.

Bosch, M., and Manacorda, M. (2012). *Social Policies and Labor Market Outcomes in Latin America and the Caribbean: A Review of the Existing Evidence* (CEP Occasional Papers 32). Center for Economic Performance.

Bosch, M., and Guajardo, J. (2012). *Labor Market Impacts of Non-Contributory Pensions: The Case of Argentina's Moratorium* (Working Paper 4094). Inter-American Development Bank.

Bosch, M., and Esteban-Pretel, J. (2013). *Labor Market Effects of Introducing Unemployment Benefits in an Economy with High Informality* (Working Paper Series 402). Inter-American Development Bank.

Bosch, M., Melguizo, A., and C. Pagés. (2013). *Better Pensions, Better Jobs: Towards Universal Coverage in Latin America and the Caribbean*. Inter-American Development Bank.

Bosch, M., and Campos-Vázquez, R. (2014). The Trade-offs of Welfare Policies in Labor Markets with Informal Jobs: The Case of the "Seguro Popular" Program in Mexico. *American Economic Journal: Economic Policy, 6*(4), 71-99.

Bosch, M., Cobacho, M.B., and Pagés, C. (2014). Effects of Non-Contributory Systems of Informality: Taking Stock of Eight Years of Implementation of Mexico's Seguro Popular. In M. Frolich, D. Kaplan, C. Pagés, J. Rigolini and D. Robalino (Eds.), *Social Insurance, Informality and Labor Markets.* Oxford University Press.

Bosch, M., Berstein, S., Castellani, F., Oliveri, M.L., and Villa, J.M. (2015). *Diagnóstico del sistema previsional colombiano y opciones de reforma* [An Assessment of the Colombian Social Security System and Reform Options] (Technical Note 825). Inter-American Development Bank.

Bosch, M., and Schady, N. (2019). The Effect of Welfare Payments on Work: Regression Discontinuity Evidence from Ecuador. *Journal of Development Economics*, *139*, 17-27.

Borrella-Mas, M.A., Bosch, M., and Sartarelli, M. (2019). *Heterogeneous Effects of a Non-contributory Pension: Evidence from Bolivia* (Working Paper 2019-35). Instituto Complutense de Análisis Económico, Universidad Complutense de Madrid.

Bozio, A., Breda, T., and Grenet, J. (2020). *Does Tax-Benefit Linkage Matter for the Incidence of Social Security Contributions?* HAL Id: halshs-02191315.

Busso, M., Madrigal, L., and Pagés, C. (2013). Productivity and Resource Misallocation in Latin America. *B.E. Journal of Macroeconomics*, *13*(1), 903-932.

CAF. (2018). *Instituciones para la productividad: hacia un mejor entorno empresarial* [Institutions for Productivity: Toward a Better Business Environment]. Corporación Andina de Fomento.

CAF. (2020a). *Los sistemas de pensiones y salud en América Latina: los desafíos del envejecimiento, el cambio tecnológico y la informalidad* [Pension and Health Systems in Latin America: Challenges of Aging, Technological Change and Informality]. Corporación Andina de Fomento.

CAF. (2020b). *El desafío del desarrollo en América Latina: políticas para una región más productiva, integrada e inclusiva* [The Challenge of Development in Latin America: Policies for a More Productive Region]. Corporación Andina de Fomento.

Cafagna, G., Aranco, N., Ibarrarán, P., Oliveri, M.L., Medellín, N., and Stampini, M. (2019). *Envejecer con cuidado. Atención a la dependencia en América Latina y el Caribe* [Aging with Care. Care for Dependent Adults in Latin America and the Caribbean]. Inter-American Development Bank.

Camacho, A., Conover, E., and Hoyos Suárez, A. (2014). Effects of Colombia's Social Protection System on Workers' Choice between Formal and Informal Employment. *World Bank Economic Review*, *28*(3), 446-466.

Card, D., and Krueger, A. (2015). *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton University Press.

Carneiro, F. (2004). Are Minimum Wages to Blame for Informality in the Labor Market? *Empirica*, *31*, 295-306.

Castañón, V., and Ferreira, O.F. (2017). Densidades de cotización en el sistema de ahorro para el retiro en México [Contribution Density in the Retirement Savings System in Mexico]. *Center for Latin American Monetary Studies Bulletin, 3,* 213-35.

Causa, O., and Hermansen, M. (2019). *Income Redistribution Through Taxes and Transfers Across OECD Countries* (Working Paper No. 1453). OECD Economics Department.

Commitment to Equity Institute. (2020). CEQ Standard Indicators Web Version 4.0. CEQData Center on Fiscal Redistribution, Commitment to Equity Institute, Tulane University.

Cooper, C., and Cuba, E. (2019). *Reforma para la reducción de la informalidad laboral y el aumento de la productividad* [Reform to Reduce Labor Informality and Increase Productivity]. Lima Chamber of Commerce.

Correa, A., García de Alba, M., Martínez, A., Martínez, C., and Sulmont, A. (2021). El sistema de protección social y laboral en México: contribución paradójica al combate a la desigualdad [The Social and Labor Protection System in Mexico: Paradoxical Contribution to the Fight against Inequality] (UNDP LAC Working Paper No. 23). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Cruces, G., Galiani, S., and Kidyba, S. (2010). Payroll Taxes, Wages and Employment: Identification Through Policy Changes. *Labour Economics*, *17*(4), 710-722.

Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., and Viollaz, M. (2011). *Intra-generational Mobility and Repeated Cross-sections. A Three-Country Validation Exercise* (Policy Research Working Paper 5916). World Bank.

Cruces, G,. and Bérgolo, M. (2013). Informality and Contributory and Non-contributory Programmes: Recent Reforms of the Social Protection System in Uruguay. *Development Policy Review*, *31*(5), 531-551.

Cruces, G,. and Gasparini, L. (2009). *Los determinantes de los cambios en la desigualdad de ingresos en Argentina: evidencia y temas pendientes* [Determinants to Changes in Income Inequality in Argentina: Evidence and Outstanding Issues] (Social Policy Working Document Series No. 5). World Bank.

Cruces, G., and Gasparini, L. (2013). *Políticas sociales para la reducción de la pobreza y la desigualdad en América Latina y el Caribe: diagnóstico, propuestas y proyecciones en base a la experiencia reciente* [Social Policies to Reduce Poverty and Inequality in Latin America and the Caribbean: Assessment, Proposals and Projections based on Recent Experience] (Working Paper 142). Centro de Estudios Distributivos, Laborales y Sociales, Universidad Nacional de la Plata.

Cuesta, J., and Olivera, M. (2014). The Impact of Social Security Reform on the Labor Market: The Case of Colombia. *Journal of Policy Modelling*, *36*(6), 1118-1134.

Da-Rocha, J., Tavares, M., and Restuccia, D. (2016). *Firing Costs, Misallocation and Aggregate Productivity* (Working Paper 23008). National Bureau of Economic Research.

Dávila, E., and Levy, S. (2003). Taxing for Equity: A Proposal to Reform Mexico's Value Added Tax. In J.A. González, V. Corbo, A. Krueger and A. Tornell (Eds.), *Latin American Economic Reform, The Second Stage*. University of Chicago Press.

De Brauw, A., Gilligan, D.O., Hoddinott, J., and Roy, S. (2015a). Bolsa Família and Household Labor Supply. *Economic Development and Cultural Change*, *63*(3), 423-457.

De Brauw, A., Gilligan, D.O., Hoddinott, J., and Roy, S. (2015b). The Impact of Bolsa Família on Schooling. *World Development*, *70*, 303-316.

De la Torre, A., and Ize, A. (2020). Accounting for Latin American Growth: A Trade and Macro Perspective. *Economia,* Fall 2020.

De Paula, A., and Scheinkman, J. (2010). Value Added Taxes, Chain Effects and Informality. *American Economic Journal: Macroeconomics*, *2*(4), 195-221.

De Soto, H. (1989). *The Other Path: The Invisible Revolution in the Third World*. Harper and Row.

Doruk, C., Dube, A., Lindner, A., and Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, *134*(3).

Doubova, S.V., García-Saisó, S., Pérez-Cuevas, R., Sarabia-González, O., Pacheco-Estrello, P., Leslie, H.H., Santamaría, C., Torres-Arreola, L.D.P., and Infante-Castañeda, C. (2018). Barriers and Opportunities to Improve the Foundations for High-quality Healthcare in the Mexican Health System. *Health Policy and Planning*, *33*(10), 1073-1082.

Dougherty, S. (2014). Legal Reform, Contract Enforcement and Firm Size in Mexico. *Review of International Economics*, *22*(4), 825-844.

Dube, A. (2019). *Impacts of Minimum Wages: Review of the International Evidence*. An Independent Report on the Impacts of Minimum Wages to Inform the UK Government's Decisions on the Remit of the Low Pay Commission beyond 2020. HM Treasury and Department for Business, Energy and Industrial Strategy.

Economic Commission for Latin America and the Caribbean. (2020). *Fiscal Panorama of Latin America and the Caribbean*. ECLAC.

Erosa, A. (2001). Financial Intermediation and Occupational Choice in Development. *Review of Economic Dynamics*, *4*, 303-333.

Escudero, V., Kluve, J., López, E., and Pignatti, C. (2019). Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta-Analysis. *The Journal of Development Studies*, *55*(12), 2644-2661.

Eslava, M., Haltiwanger, J., and Pinzón, A. (2019). *Job Creation in Colombia vs. the US: 'Up or Out Dynamics' Meets 'the Life Cycle of Plants'* (Working Paper 25,550). National Bureau of Economic Research.

Fernández-Arias, E. (2017a). *Productivity and Factor Accumulation in Latin America and the Caribbean: A Data Base.* Research Department, Inter-American Development Bank.

Fernández-Arias, E. (2017b). *Productivity and Factor Accumulation in Latin America and the Caribbean: Database Description*. Research Department, Inter-American Development Bank.

Fernández-Arias, E., and Fernández-Arias, N. (2021). *The Latin American Growth Shortfall: Productivity and Inequality*. Background paper for the 2021 Report of Latin American and Caribbean Bureau of the UNDP.

Fernández, C., and Villar, L. (2017). The Impact of Lowering the Payroll Tax on Informality in Colombia. *Economia Journal*, Fall, 125-155.

Ferreira, F., and Robalino, D. (2011). Social Protection in Latin America: Achievements and Limitations. In J.A. Ocampo and J. Ros (Eds.), *The Oxford Handbook of Latin American Economics*. Oxford University Press.

Ferreira, F., Firpo, S., and Messina, J. (2021). Trade, Labor Market Experience and Falling Earnings Inequality in Brazil: 1995-2012. *World Bank Economic Review,* forthcoming.

Fiszbein, A., and Schady, N. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. World Bank.

Firpo, S., and Portella, A. (2021). *Informal and Small: How Labor Market Regulations, the Social Protection System and Special Tax Regimes Impact Formalization, Inequality, Income Volatility, Firm Size and Labor Productivity in Brazil* (UNDP LAC Working Paper No. 22). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Flabbi, L. (2021). *Implications of Minimum Wage Policies in Labor Markets with High Informality and Frictions* (UNDP LAC Working Paper No. 25). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Flinn, C.J. (2011). *The Minimum Wage and Labor Market Outcomes*. MIT Press.

Flores, I., Sanhueza, C., Atria, J., and Mayer, R. (2019). Top Incomes in Chile: A Historical Perspective on Income Inequality, 1964-2017. *Review of Income and Wealth*, *66*(4), 850-874.

Galiani, S., and Weinschelbaum, F. (2012). Modeling Informality Formally: Households and Firms. *Economic Inquiry*, *50*(3), 821-838.

Galiani, S., Gertler, P., and Bando, R. (2016). Non-Contributory Pensions. *Labour Economics, 38*(C), 47-58.

Garganta, S., and Gasparini, L. (2015). The Impact of a Social Program on Labor Informality: The Case of AUH in Argentina. *Journal of Development Economics*, *115*, 99-110.

Garganta, S., Gasparini, L., and Marchioni, M. (2017). Cash Transfers and Female Labor Force Participation: The Case of AUH in Argentina. *IZA Journal of Labor Policy*, *6*(10), 1-22.

Gasparini, L., Haimovich, F., and Olivieri, S. (2009). Labor Informality Bias of a Poverty-Alleviation Program in Argentina. *Journal of Applied Economics*, *12*(2), 181-205.

Gerard, F., Naritomi, J., and Seibold, A. (2018). *Tax Systems and Inter-Firm Trade: Evidence from the VAT in Brazil*. Mimeo.

Gruber, J. (1997). The Incidence of Payroll Taxation: Evidence from Chile. *Journal of Labor Economics*, *15*(53), 72-101.

Grushka, C. (2019). *The Within-system Redistribution of Contributory Pensions Systems: a Conceptual Framework and Empirical Method of Estimation* (No. 91). Department of Economics, Tulane University.

Guerrero, I., López-Calva, L.F., and Walton, M. (2009). The Inequality Trap and Its Links to Low Growth in Mexico. In S. Levy and M. Walton (Eds.), *No Growth without Equity? Inequality, Interests and Competition in Mexico*. Palgrave MacMillan for the World Bank.

Guha-Khasnobis, B., Kanbur, R., and Ostrom, E. (2006). *Linking the Formal and Informal Economy: Concepts and Policies*. Oxford University Press.

Guner, N., Ventura, G., and Xy, Y. (2008). Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics*, 11, 721-44.

Ham, A. (2018). The Consequences of Legal Minimum Wages in Honduras. *World Development*, 102, 135-157.

Ham, A., and Membreño, S. (2021). ¿Cuán *efectiva es la protección social en Honduras?* [How Effective Is Social Protection in Honduras? (UNDP LAC Working Paper No. 21). Background paper for the United Nations Development Programme 2021 Regional Human Development Report for Latin America and the Caribbean.

Hausmann, R. (2011). Structural Transformation and Economic Growth in Latin America. In J.A. Ocampo and J. Ros (Eds.), *The Oxford Handbook of Latin American Economics*. Oxford University Press.

Heckman, J., and Pagés, C. (2004). *Law and Employment: Lessons from Latin America and the Caribbean*. University of Chicago Press.

Hernández, G. (2012). Payroll Taxes and the Labor Market: A Computable General Equilibrium Analysis. *Latin American Journal of Economics*, *49*, 99-123.

Hirschman, A.O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Harvard University Press.

Hoynes, H., and Patel, A. (2018). Effective Policy for Reducing Poverty and Inequality? The Earned Income Tax Credit and the Distribution of Income. *Journal of Human Resources, 53*(4), 859-890.

Hsieh, C., and Klenow, P. (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics, 124*(4), 1043-1448.

Inter-American Development Bank. (2010). *The Age of Productivity: Transforming Economies from the Bottom Up.* Development in the Americas Report.

Inter-American Development Bank. (2014). *Rethinking Productive Development: Sound Policiesand Institutions for Economic Transformation.* Development in the Americas Report.

Inter-American Development Bank. (2017a). *Learning Better: Public Policy for Skills Development.* Development in the Americas Report. Washington, DC.

Marchán, E., Espinasa, R., and Yépez, A. (Eds.). (2017b). *The Other Side of the Boom: Energy Prices and Subsidies in Latin America and the Caribbean during the Super-Cycle. Inter-American Development Bank.*

Cavallo, E., and Powell, A. (Eds.). (2019). *Building Opportunities for Growth in a Challenging World*. 2019 Macroeconomic Report. Inter-American Development Bank.

Busso, M., and Messina, J. (Eds.). (2020a). *The Inequality Crisis: Latin America and the Caribbean at the Crossroads*. Inter-American Development Bank.

Cavallo, E., Serebrisky, T., and Powell, A. (Eds.). (2020b). *From Structures to Services: The Path to Better Infrastructure in Latin America and the Caribbean.* Inter-American Development Bank.

Izquierdo, A., Altamirano, M., Alaimo, V., and Dborkin, D. (2018). *Subsidizing Informality? Non-Contributory Public Spending in Latin America and the Caribbean* (Technical Note IDB-TN-1540). Inter-American Development Bank.

Kanbur, R., and Keen, M. (2014). *Thresholds, Informality and Partitions of Compliance* (Working Paper 2014-11). School of Applied Economics and Management, Cornell University.

Kaplan, D., Sadka, J., and Silva-Méndez, J.L. (2008). Litigation and Settlement: New Evidence from Labor Courts in Mexico. *Journal of Empirical Legal Studies*, *5*(2), 309-350.

Kaplan, D., and Sadka, J. (2011). *The Plaintiff's Role in Enforcing a Court Ruling: Evidence from a Labor Court in Mexico* (Working Paper 264). Inter-American Development Bank.

Kaplan, D., and Levy, S. (2014). The Evolution of Social Security Systems in Latin America, in M. Frolich, D. Kaplan, C. Pagés, J. Rigolini and D. Robalino (Eds.), *Social Insurance, Informality and Labor Markets*, Oxford University Press.

Katz, L., and Murphy, K. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, *107*(1), 35-78.

Kluve, J. (2016). *A Review of the Effectiveness of Active labor Market Programmes with a Focus on Latin America and the Caribbean* (Working Paper 9). Research Department, International Labor Office.

Kugler, A. (2005). Wage Shifting Effects of Severance Payments Savings Accounts in Colombia. *Journal of Public Economics*, 89, 487-500.

Kugler, A., and Kugler, M. (2009). Labor Market Effects of Payroll Taxes in Developing Countries: Evidence from Colombia. *Economic Development and Cultural Change*, *57*(2), 335-358.

Kugler, A., Kugler, M., and Herrera, L. (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform. *Economia*, Fall, 3-40.

Laeven, L., and Woodruff, C. (2007). The Quality of the Legal System, Firm Ownership and Firm Size. *Review of Economics and Statistics*, *89*(4), 601-614.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2017). Life-Cycle Wage Growth Across Countries, *Journal of Political Economy*, forthcoming.

Lagomarsino, A., and Nocetto, L. (2020). Do Cash Assistance Programs Create Welfare Traps?, mimeo, Economics Department, Harvard University

Leal, J. (2014). Tax Collection, the Informal Sector and Productivity. *Review of Economic Dynamics*, *17*(2), 262-286.

Levy, S. (2006). *Progress Against Poverty: Sustaining Mexico's Progresa-Oportunidades Program*. Brookings Institution Press.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality and Economic Growth in Mexico*. Brookings Institution Press.

Levy, S. (2009). Social Security Reform in Mexico: For Whom? In Levy, S. and Walton, M. (Eds.), *No Growth without Equity? Inequality, Interests and Competition in Mexico.* Palgrave MacMillan for the World Bank, New York.

Levy, S. (2018). *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Inter-American Development Bank.

Levy, S. (2019). *Una prosperidad compartida: transformando la seguridad social en México para crecer con equidad* (Working Paper 1788). Inter-American Development Bank.

Levy, S. (2020). *Suggestions for the Emergency* (Policy Document Series No. 2). United Nations Development Program, Latin American and the Caribbean.

Levy, S., and López-Calva, L.F. (2020). Persistent Misallocation and the Returns to Education in Mexico. *World Bank Economic Review*, *34*(2), 284-311.

López-Calva, L.F., and Ortiz-Juárez, E. (2013). A Vulnerability Approach to the Definition of Middle Class. *Journal of Economic Inequality, 12(*1), 23-47.

López-Calva, L.F., Cruces, G., Lach, S., and Ortiz-Juárez, E. (2014). Clases medias y vulnerabilidad a la pobreza: reflexiones desde América Latina. *El trimestre económico, 81*(2), 281-307.

López-Calva, L.F, and Lustig, N. (2010). *Declining Inequality in Latin America: A Decade of Progress?* Brookings Institution and UNDP.

Lustig, N. (2020). Desigualdad y política social en América Latina. *El desafío del desarrollo en América Latina: políticas para una región más productiva, integrada e inclusiva*. Corporación Andina de Fomento.

Medeiros, M., Britto, T., and Veras Soares, F. (2008). *Targeted Cash Transfer Programs in Brazil: BPC and Bolsa Família* (Working Paper 46). International Policy Centre for Inclusive Growth.

Messina, J., and Silva, J. (2019). Twenty Years of Wage Inequality in Latin America. *World Bank Economic Review*, pp. 1-31.

Molina Millán, T., Barham, T., Macours, K., Maluccio, J.A., and Stampini, M. (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence. *The World Bank Research Observer, 34*.

Molina Millán, T., Macours, K., Maluccio, J.A., and Tejerina, L. (2020). Experimental Long-Term Effects of Early-Childhood and School-Age Exposure to a Conditional Cash Transfer Program. *Journal of Development Economics*, *143*.

Morales, M., and Olate, C. (2021). ¿Qué *tan eficaz es la protección social en Chile?* [How Effective is Social Protection in Chile? (UNDP LAC Working Paper No. 15). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Neumark, D., and Shirley, P. (2021). *Myth or Measurement: What Does the New Minimum Wage Research Say about Minimum Wages and Job Loss in the United States* (Working Paper 28388). National Bureau of Economic Research.

Ñopo, H. (2021). *Políticas de protección social y laboral en el Perú: una espiral de buenas intenciones, malos resultados y peores respuestas* [Social Protection and Labor Policies in Peru: A Spiral of Good Intentions, Bad Outcomes and Worse Responses] (UNDP LAC Working Paper No. 17). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Organization for Economic Co-operation and Development (2017). *Employment and Labour Market Statistics*. http://dx.doi.org/10.1787/data00306.

Organization for Economic Co-operation and Development et al. (2020). *Revenues Statistics in Latin America and the Caribbean 2020*. OECD Publications.

Ospina, M., and Saavedra, F. (2013). Social Assistance and Informality: Examining the Link in Colombia. *Revista de economía del Rosario*, *21*(1), 81-120.

Pagés, C. (2017). Do Payroll Tax Cuts Boost Formal Jobs in Developing Countries? *IZA World of Labor*, *345*, 1-9.

Parker, S., and Petra, Todd. (2017). Conditional Cash Transfers: The Case of Progresa-Oportunidades. *Journal of Economic Literature*, *55*(3), 866-915.

Peña, A., and Ñopo, H. (2021). *Políticas de protección social y laboral en Ecuador* [Social Protection and Labor Policies in Ecuador] (UNDP LAC Working Paper No. 19). Background paper for the United Nations Development Programme, Regional Human Development Report for Latin America and the Caribbean.

Prada, M.F., Rucci, G., and Urzua, S. (2015). *The Effects of Mandated Childcare on Female Wages in Chile* (Working Paper 594). Inter-American Development Bank.

Quintin, E. (2008a). Contract Enforcement and the Size of the Informal Economy. *Economic Theory*, *37,* 395-416.

Quintin, E. (2008b). Limited Enforcement and the Organization of Production. *Journal of Macroeconomics, 30*, 1222-1245.

Restuccia, D., and Rogerson, R. (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics*, *11*(4), 707-720.

Rocha, R., Ulyssea, G., and Rachter, L. (2018). Do Lower Taxes Reduce Informality? Evidence from Brazil. *Journal of Development Economics*, *134*, 28-49.

Rodríguez, C., López-Calva, L., and Barriga, O. (2020). *The Effects of Local Market Concentration on International Competition and Firm Productivity*, World Bank Policy Research Working Paper 9210, Washington, DC.

Rofman, R. (2021). *Los regímenes previsionales de excepción en Argentina* (Working Paper, Social Protection Program). CIPPEC.

Ronconi, L. (2021). *How do Labor Unions Affect Inequality and Efficiency in Latin America and the Caribbean?* Department of Economics, University of Buenos Aires.

Scott, J. (2014). Redistributive Impact and Efficiency of Mexico's Fiscal System. *Public Finance Review*, *42*(3), 368-90.

Scott, J., de la Rosa, E., and Aranda, R. (2017). *Inequality and Fiscal Redistribution in Mexico: 1992 2015* (WIDER Working paper 2017/194). World Institute for Development Economics Research.

Skoufias, E., and Maro, V.D. (2008). Conditional Cash Transfers, Adult Work Incentives, and Poverty. *The Journal of Development Studies*, *44*(7), 935-960.

Skoufias, E., Unar, M., and de Cossio, T.G. (2013). The Poverty Impacts of Cash and In-kind Transfers: Experimental Evidence from Rural Mexico. *Journal of Development Effectiveness*, *5*(4), 401-429.

Soares, F.V., Ribas, R.P., and Osório, R.G. (2010). Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programs in Comparative Perspective. *Latin American Research Review*, *45*(2), 173-90.

Summers, L. (1989). Some Simple Economics of Mandate Benefits. *American Economic Review*, *79*(2), 177-83.

Ulyssea, G. (2018). Firms, Informality and Development: Theory and Evidence from Brazil. *American Economic Review*, *108*(8), 2015-47.

Ulyssea, G. (2020a). *Formal and Informal Firm Dynamics*. Department of Economics, Oxford University.

Ulyssea, G. (2020b). Informality: Causes and Consequences for Development. *Annual Review of Economics*, *12*(1), 525-546.

Uthoff, A. (2011). Social Security Reforms in Latin America, in J.A. Ocampo and J. Ros (Eds.), *The Oxford Handbook of Latin American Economics*. Oxford University Press.

Vakis, R., Rigolini, J., and Lucchetti, L. (2016). *Left Behind: Chronic Poverty in Latin America and the Caribbean*. World Bank.

Vázquez, G., and Maurizio, R. (2015). Argentina: Impacts of the Child Allowance Programme on the Labour-Market Behaviour of Adults. *CEPAL Review*, pp. 115-137.

World Bank. (2014). *Latin American Entrepreneurs: Many Firms but Little Innovation*. Office of the Chief Economist, Latin American and the Caribbean.

COMPLETE LIST OF BACKGROUND PAPERS OF THE 2021 UNDP LAC REGIONAL HUMAN DEVELOPMENT REPORT

Working paper 01
The Changing Picture of Inequality in Latin America: Evidence for Three Decades
Leonardo Gasparini / Guillermo Cruces

Working paper 02
Intergenerational Mobility in Latin America in Comparative Perspective
Florencia Torche

Working paper 03
Longer-run Perspectives on Intergenerational Mobility in Latin America: 1933-1992
Guido Neidhoeffer

Working paper 04
The Latin American Growth Shortfall: Productivity and Inequality
Eduardo Fernández Arias / Nicolás Fernández Arias

Working paper 05
Big Business and the Inequality Trap in Latin America: Taxes, Collusion, and Undue Influence
Ben Ross Schneider

Working paper 06
Lobbying for Inequality? How Business Elites Shape Policy in El Salvador, Honduras, and Guatemala
Mónica Pachón / Javier Broló

Working paper 07
How do Labor Unions affect Inequality and Efficiency in Latin America and the Caribbean?
Lucas Ronconi

Working paper 08
The Political Economy of The Telecommunications Reform in Mexico
Alejandro Castañeda / Daniel Ruiz

Working paper 09
El impacto de las instituciones del mercado laboral sobre "markups" y "markdowns": evidencia de los sectores manufacturero y de servicios en Uruguay
Nestor Gandelman

Working paper 10
Labor-Market Concentration and Earnings: Evidence from Chile
Alvaro Garcia

Working paper 11
Poder de mercado, desigualdad y productividad en Latinoamérica
Marcela Eslava / Marcela Melendez / Nicolás Urdaneta

Working paper 12
The Effects of Violence on Inequality in Latin America and the Caribbean: A Research Agenda
Ana Arjona

Working paper 13
Inequality and Crime in Latin America and the Caribbean: New Data for an Old Question
Ernesto Shargrodsky

Working paper 14
El sistema de protección social en Colombia
Colombia: Marcela Meléndez / Francisco Alvarado / Marcela Pantoja

COMPLETE LIST OF BACKGROUND PAPERS OF THE 2021 UNDP LAC REGIONAL HUMAN DEVELOPMENT REPORT

Working paper 15
¿Qué tan eficaz es la Protección Social en Chile?
Chile: Camila Olate / Matías Morales

Working paper 16
Proteccion social, formalidad y subsidios cruzados: evidencia para Argentina
Argentina: Maria Laura Alzua

Working paper 17
Políticas de protección social y laboral en el Perú. Una espiral de buenas intenciones, malos resultados y peores respuestas
Perú: Hugo Ñopo

Working paper 18
Políticas de protección social y laboral en la República Dominicana
Rep. Dom: Hugo Ñopo / Socrates Barinas

Working paper 19
Políticas de protección social y laboral en Ecuador
Ecuador: Hugo Ñopo / Alejandra Peña

Working paper 20
The persistence of inequality and low growth in Latin America and the Caribbean
Jamaica: Manuel Mera

Working paper 21
¿Cuán efectiva es la protección social en Honduras?
Honduras: Andres Ham / Sergio Membreño

Working paper 22
Informal and small: How labor market regulations, the social protection system and special tax regimes impact formalization, inequality, income volatility, firm size and labor productivity in Brazil
Brasil: Sergio Firpo

Working paper 23
El sistema de protección social en México y su contribución a combatir las desigualdades
México: Annabelle Sulmont / Cynthia Martínez

Working paper 24
Hora de cambiar el rumbo: Un ensayo sobre la Protección Social en Latinoamérica
Conceptual: Santiago Levy

Working paper 25
Implications of Minimum Wages Policies in Labor Markets with High Informality and Frictions
Salario Mínimo. Luca Flabbi