The End of Informality in Mexico?
Fiscal Reform for Universal Social Insurance

Arturo Antón, Fausto Hernández and Santiago Levy


1 Antón and Hernández are with the Centro de Investigación y Docencia Económicas (CIDE); Levy with the Inter-American Development Bank (IADB). The views expressed here are the author’s and do not necessarily coincide with those of the institutions they are affiliated with. Antón and Hernández work was financed by the IADB. We thank Victoria Fazio and Carlos Herrera for excellent collaboration and Emma Aguila, Mariano Bosch, Ana Corbacho, Guillermo Cruces, Carlos Elizondo, Eduardo Lora, Teresa Ter-Minassian and Michael Walton for very useful suggestions.
Abstract

Mexico is characterized by a dual social insurance architecture. Firms and workers in salaried contractual relations are obligated to pay for a bundled set of health, pension and related programs. Non-salaried workers benefit from an unbundled set of parallel programs paid by the government. We develop a model to study the implications of this architecture in a context of informality and imperfect tax enforcement. We argue that this architecture: (i) provides workers with erratic and incomplete coverage against risks, (ii) fosters evasion and narrows the tax base, (iii) delinks contributions from benefits undermining fiscal sustainability, and (iv) distorts the labor market lowering real wages and total factor productivity.

We propose a reform to shift taxation for social insurance from labor to consumption. We show that by setting a uniform value added tax rate of 16 percent it is possible to provide all workers with the same health and pension benefits and Hicks-compensate poor households for the VAT increase at a net fiscal cost of 0.3 percent of GDP. We argue that our proposal: (i) effectively protects all workers against risks, (ii) reduces distortions in the labor market stemming from social insurance tax-cum-subsidies allowing for an increase in the real wage despite the higher VAT, (iii) raises total factor productivity, (iii) contributes to reduce poverty and income inequality, (iv) links contributions with benefits ensuring fiscal sustainability, (v) increases aggregate savings for retirement, and (vi) reduces evasion and widens the tax base.
1. Introduction

On November 17, 1881, Germany’s Emperor William the First delivered an Imperial Message to the National Parliament (Reichstag), expressing his conviction that “the healing of social wrongs must be sought not solely through the repression of social democratic excesses but just as much by positively advancing the well-being of workers”. Soon after that, under the leadership of his Chancellor Otto von Bismarck, the Reichstag approved the Law Concerning Health Insurance for Workers (1883), the Accident Insurance Act (1884) and the Law on Invalidity and Old Age Insurance for Workers, Journeymen and Apprentices (1889). Thus was social insurance born, the cornerstone of the modern Welfare State.

On December 1, 1940, Mexico’s President Avila Camacho delivered his Inaugural Address to Congress, expressing his conviction that “we should all assume the purpose, to which I shall display all my energies, that soon social security laws protect all Mexicans in times of adversity, during orphanage, when women are widowed, in sickness, unemployment and old age, to substitute this secular situation that as a result of the poverty of the Nation we have all lived with”. Soon after that, in January 1943, Congress approved the country’s first Social Security Law. Thus was social insurance born in Mexico.

Bismarck’s social insurance model was based on a key idea: benefits for workers are paid by the firms hiring them through an ear-marked tax (or contribution) proportional to workers’ wages (initially set at six per cent!). As a result, the adjective contributory has been added to this model of social insurance (henceforth CSI). Why was this method of financing chosen? Two motivations stand out. The first, which we call the administrative convenience motive, is that wages can be measured more easily by the fiscal authorities than other sources of income; as a result, compliance can be exercised by auditing firms’ payrolls and by the fact that firms have a fixed location where assets can be impounded. The obligatory nature of CSI —critical from the point of view of risk pooling— can thus be enforced. The second, which we call the redistributive motive, is that it can be argued that through CSI a share of firm’s profits is re-distributed to workers. These are clearly important motivations. But Bismarck’s idea has a fatal flaw: by design, the coverage of social insurance is limited by the method of financing. While all workers are exposed to the risks being insured, even under full compliance only workers hired by firms and receiving payments in the form of wages are actually insured; self-employed workers and those engaged with firms with non-salaried contracts are left-out.
Administrative convenience and redistribution motivations loom large in Mexico’s case because of its high income inequality and because its State has historically been characterized by its weak capacity to tax. In this context, a tax that could be easily collected and also served to reduce inequality was attractive indeed. Thus, the fatal flaw in Bismarck’s social insurance model was replicated in Mexico’s model. Despite President’s Avila Camacho purpose that “…soon social security laws protect all Mexicans…”, coverage was limited to salaried workers (and their families). As a result, almost 70 years after the country’s Social Security Law was enacted, more than half of Mexican workers are still excluded. Mexico’s Welfare State was born “truncated”, has remained so since then and, under present course, will remain so forever.

Non-salaried workers represent a major challenge for social insurance (henceforth, SI). Under CSI there is no possibility of extending coverage to them precisely because they are non-salaried. But they need insurance nonetheless: they can get sick, die suddenly, suffer from disability, or fail to save for retirement as much as salaried workers can. As a result, over the last seven decades the Mexican government has gradually created programs to provide non-salaried workers with at least a subset of the benefits that salaried workers receive through CSI. These efforts began with low quality health programs (relative to those provided by CSI). But as years have gone by, and particularly since the mid 1990s, these health programs, and new ones in pensions, day care and housing, have expanded markedly in scope, quality, budget and coverage; see Levy (2008). In 2008, the government spent at least 1.25% of GDP subsidizing health, housing, pension and day care programs for those excluded from CSI; this contrasts with subsidies of 0.5% of GDP for CSI. Wily nilly, a parallel system of SI is being created in Mexico, financed from revenues other than wage taxes. Following standard practice, we label this parallel system non-contributory social insurance, NCSI.

2 Mexico has designed schemes to promote the affiliation of non-salaried workers to CSI, but they have failed. In 2006 only 0.5 million out of a potential universe of 25 million of workers were enrolled in such a scheme called Voluntary Affiliation into the Obligatory Regime (Levy, 2008). They key difficulty, of course, is its voluntary nature. But even if the Law obligated non-salaried workers to enroll, there would be no enforcement mechanism as observing their earnings is difficult, and even if they could be observed, collection and enforcement costs would make this prohibitive. One can tax the wage bill of a firm with 100 workers in a fixed location with relative ease, and credibly impound the firm’s assets if it does not comply; it is a different matter to individually tax the earnings (out of wages? capital?) of 100 self-employed workers dispersed through many locations, and to impound assets that may be worth less than the taxes due, or the costs of collection.

3 Over the period 1998-2007 subsidies for NCSI programs grew by 110%. In February 2006 President Fox issued a Decree creating the National Council for Social Protection to “…ensure the functional integration of benefits in health, housing and savings for retirement, among others, that the Federal Government offers to the population lacking social security coverage”; see Levy (2008).
Elsewhere, one of us has argued that the CSI + NCSI configuration is bad social policy and bad economic policy (Levy, 2008). It is bad social policy because, first, during their life cycle workers transit between salaried and non-salaried status, implying that at times they receive SI through CSI and at times through NCSI: sometimes they are protected against some risks, and sometimes not; sometimes coverage is obligatory and sometimes voluntary. And second, because CSI taxes are an ineffective redistributive tool as their incidence falls mainly on workers (the letter and intention of the Social Security Law notwithstanding). As a result, neither the insurance nor the redistributive objectives are effectively reached. On the other hand, it is bad economic policy because the CSI + NCSI configuration translates into a tax on salaried labor and a subsidy to non-salaried labor. Firms and workers naturally react to this tax-cum-subsidy by shifting into activities intensive in non-salaried labor, and when the production technology limits this, by evading the tax on salaried labor adjusting firm size, duration of contracts and other variables, resulting in factor misallocations that are costly from the point of view of productivity and growth (Busso et al., 2011).

In this context, Levy (2008) has argued that there is a better alternative to the CSI + NCSI configuration: providing all workers with the same SI regardless of their salaried or non-salaried status, by changing the mix of financing from the current blend of wage taxes and other sources of revenues, to a single source based on an ear-marked consumption tax. The proposal rests on a simple idea: by changing the point where SI contributions are collected from “the door of the factory” to “the door of the store”, the fatal flaw of Bismarck’s idea can be overcome since the second door, as opposed to the first, does not discriminate between salaried and non-salaried workers; contributions “at the door of the store” cannot be evaded by modifying the nature of the labor contract or its duration, or the legal status of the firm.

---

4 Of course, there is no such thing as non-contributory social insurance, as benefits need to be paid by someone. More precise expressions are “social insurance contributory from wages” and “social insurance contributory from general sources of revenues”. However, to avoid confusion we follow standard practice and use the contributory and non-contributory labels. Importantly, NCSI programs should not be confused with targeted programs for the poor that transfer income through various means including conditional cash transfer programs like Progresa-Oportunidades (sometimes labeled social assistance programs); NCSI programs target beneficiaries on the basis of their labor status, not income levels. Also, to avoid even more confusion we eschew the expressions “social protection” or “social safety nets”.

5 This is an inappropriate design from an insurance perspective as it reduces the scope for risk pooling, creates adverse selection problems and induces moral hazard behavior. It also seriously limits the effectiveness of retirement pensions (regardless of whether they are defined benefit or defined contribution), as workers accumulate for their pensions only during the fraction of their working lives that they have a salaried job with a firm that complies with the Law.
More particularly, the proposal recognizes that because CSI taxes reduce salaried workers’ wages, they basically change the composition of their consumption: less free disposable income in exchange for SI benefits. Differently put, CSI taxes are equivalent to a consumption tax on salaried workers ear-marked for SI benefits. Two questions are relevant in this context. Why are salaried workers obligated to pay directly for their SI through lower wages, while non-salaried workers get it from other sources of revenue? Would it not be more efficient to lower all workers’ disposable income through the same mechanism—a consumption tax—and pursue redistribution objectives through another mechanism? And while a consumption tax creates its own distortions (and can also be evaded), these are substantially less costly than those created by the combination of taxes on salaried labor and subsidies to non-salaried labor implicit in Mexico’s current CSI + NCSI model. A new SI architecture based on an ear-marked consumption tax allows the government to minimize distortions in the market for the most important non-traded input: labor. At the same time, critically, it allows the government to provide the same SI to all workers.

This paper studies a fiscal reform that changes wage-based to consumption-based contributions to achieve universal social insurance, USI. In particular, we provide a quantitative evaluation of Levy’s proposal to unify the differentiated structure of value added tax (VAT) rates at the level of the highest existing rate, ear-mark the additional revenues together with existing subsidies for CSI and NCSI to fund the same SI benefits for all workers, drastically reduce CSI taxes so that they only finance risks specific to salaried status and complementary pension benefits, and compensate low income households for the negative income effects of the VAT increase.6

Mexico, as other countries with a dualistic model of SI, is characterized by the co-existence of formal and informal economic activity, a phenomenon directly related to this dualism. It is also characterized by high tax evasion. Because the interplay between formality-informality, on one hand, and legality-illegality, on the other, is of the essence for the problem at hand, we develop a model where these features are prominent. To capture the core distinction created by Bismarckian SI, we assume there are two sectors with alternative contractual relations: one where firms and workers have salaried contractual relations and a second one where workers are self-employed or where firms and workers have non-salaried

---

6 Narro et al. (2011) also call for a major revision of the functioning of Mexico’s social security system, but do not dwell on its fiscal implications.
contractual relations. Workers are formal when they are covered by CSI and informal when they are covered by NCSI. In parallel, we assume enforcement is imperfect and focus on firm’s incentives to engage in illegal behavior by evading CSI or VA taxes. In this context, firms may not enroll all of their salaried workers in CSI; informal employment is thus made up of non-salaried and illegally hired salaried workers.

A central feature of our model is that the division of employment between salaried and non-salaried, on one hand, and between formal and informal, on the other, is endogenous to the taxes and subsidies implicit in the CSI + NCSI configuration and to the VAT. The same holds for the degree of firm compliance with VA and CSI taxes. Firms are price takers and maximize profits in the usual fashion, and as part of their maximizing strategies they may evade CSI or VA taxes. If firms evade they face an endogenous probability of being detected by the relevant authority, with such probability depending positively on firm’s size for the case of VAT, and on firm’s size and level of illegal salaried employment for the case of CSI taxes. These assumptions, which reflect the institutional context of Mexico where VAT is collected by the Finance Ministry and CSI taxes by the Social Security Institute, imply that the rates of tax compliance (and thus the rates of tax evasion) are specific for each tax and firm. For example, firms may optimally choose not to pay CSI taxes on some workers if the probability of detection is relatively low. In that case, firms hire some salaried workers legally and some illegally. At the same time, depending also on the probabilities of detection and on fines, firms may also partly or fully evade the VAT, and this behavior may also affect their compliance with CSI taxes; differently put, the VAT may augment informal employment. As a result, the tax base and actual revenues depend on the array of CSI-NCSI and VA taxes and subsidies, and on the extent of informality in the economy.

Our model incorporates intermediate inputs, an important feature given our focus on the VAT. Firms producing intermediate inputs can also evade. Because the VAT is assumed to work via the credit method, where firms who pay VAT on an intermediate input can receive a credit only if they pay VAT on their own output, the model allows for a mechanism where informality is transmitted across the production chain, as in de Paula and Scheinkman (2010).

---

7 This includes, for example, production in small family firms, or production in agriculture by farmers individually or in sharecropping agreements. The key point is that there are no salaried contractual relations – no subordination of a worker to a boss/firm in exchange for a wage– and therefore, according to Mexico’s Labor and Social Security laws, no obligation that workers be enrolled in CSI.
To capture how the economy responds to the fiscal-cum-social reform that we propose, we first focus on how evasion of CSI and VA taxes changes under different policy combinations, given the government’s enforcement efforts. In doing this we stress a very important point, often missed in discussions of SI: the composition of public spending in SI is as important as the level. Higher subsidies to NCSI enlarge the size of the informal sector directly, induce more evasion and lower total tax revenues; higher CSI taxes also induce more evasion and enlarge the informal sector, but increase total revenues. On the other hand, subsidies to CSI reduce the incentives to evade, increase formal employment and expand the tax base and total tax revenues. Differently put, in an economy characterized by the CSI + NCSI duality, the method of increasing public spending to expand the coverage of SI has large fiscal implications. Thus, our model captures the sharp trade-offs between social and economic objectives present in a dualistic system of SI: on one hand, attempts to complete Mexico’s “truncated” Welfare State expanding NCSI will have negative implications for fiscal sustainability (and for productivity); on the other, not attempting to do so would leave millions of workers with no or limited coverage of SI.

The key result of our paper is to show that these trades-offs can be avoided by moving to a model of USI. The proposed fiscal-cum-social reform yields important gains in workers’ real wages and welfare, without placing an undue burden on the government’s fiscal balance: in our central estimate we find that after fully compensating various groups, the fiscal deficit under USI would be 0.3% of GDP higher vis-à-vis the status quo. The effort needed to accommodate this figure by raising revenues or reducing expenditures elsewhere would need to be balanced against what the proposal implies, namely, by far the most ambitious social reform in Mexico since the birth of SI in 1943; a drastic reduction of productivity-reducing distortions in the labor market; and the medium term fiscal sustainability of SI policy.

To illustrate this key result, we calibrate our model to replicate key features of Mexican data observed in 2008 including the distribution of employment by firm size and formality status, the observed levels of CSI and VAT evasion and, critically, the government’s fiscal balance. We next proceed to evaluate our proposal in two stages. First, we model a fiscal reform that simultaneously imposes a 16 percent VAT on all goods and completely eliminates all CSI taxes and NCSI subsidies; second, we examine the same reform re-introducing taxes on salaried labor—but at drastically lower rates than under CSI--, to ensure that pension benefits for salaried workers are at least as generous as under CSI.
In the first stage we find that, after taking tax evasion by firms into account, a complete shift from labor to consumption taxes increases the VAT revenue/GDP ratio from 0.038 to 0.071. This large increase in revenues results from eliminating distortions in both labor and goods markets at the same time. These resources more than compensate lost revenues from eliminating CSI taxes, and at the same time finance the extra expenditure required to sustain USI. We estimate as well that after the VAT increase real wages would be 15 percent higher, as labor is allocated more efficiently when the tax on salaried labor and the subsidy to non-salaried labor are eliminated. In parallel, the associated contraction in non-salaried employment contributes to expand the tax base and therefore to the increase in VAT revenues (since both tax rates and tax compliance are lower for the self-employed and for family firms). Finally, we estimate that the transit from the current CSI + NCSI to the USI model implies an increase in public spending in social insurance of 2.8% of GDP, as all workers receive the same health and pension benefits. Critically, this large increase in social spending can occur while labor market distortions are substantially diminished and fiscal sustainability strengthened, as opposed to what would occur if the same expansion in SI occurred by raising the benefits provided by NCSI to the level of CSI (as has been gradually happening over the last decades).

In the second stage, to ensure that salaried workers receive at least the same pension benefits as they do under CSI, we extend our proposal complementing the VAT-financed USI benefits for all workers with additional benefits for salaried workers paid from wage-based taxes. As a result, these latter taxes would not be completely eliminated; nevertheless, we estimate that they would fall from 32% of the wage under CSI to an average of 5.5% under our proposal. This drastic reduction in the wedge between what firms pay for salaried labor and what salaried workers earn allows for both lower labor costs and higher real wages (despite the VAT increase, as noted). This reduction in firms’ labor costs is de facto equivalent to a real depreciation of approximately 10% (a result at times labeled as “fiscal devaluation”).

To avoid any regressive effect of the VAT increase on the poor, we calculate the resources needed to Hicks-compensate households in the first two deciles of the distribution. When these compensations are considered, the proposal has a large and unambiguously positive effect on reducing poverty: under USI poor Mexican workers would earn a higher real wage, would be covered by the same SI as other workers, would pay no net additional taxes,

---

8 The simultaneity of the changes in CSI and VA taxes is central to our results, in particular to the reduction of informality. As shown below, a pure increase in VA taxes, while increasing revenues, raises as well the level of informality, a result consistent with the literature; see Emran and Stiglitz (2005) and Keen (2008).
and would experience the largest reduction in wage-based taxes (from 32 to 1.5%). Equally importantly, our proposal would complement current efforts to combat poverty through direct income transfers, particularly those effected through the Progresa-Oportunidades program, with significant improvements in the functioning of the labor market and in poor workers’ possibilities of finding more stable jobs with better prospects for productivity gains through their life-cycle.

Although we think these results are interesting in their own right, the focus of the paper is on the policy implications. As a result, the robustness of our numerical estimates is critical, particularly those centered on the additional revenues produced by the VAT reform. To assess this issue, we note that the core transmission mechanism in our model is the re-allocation of labor, which shifts between non-salaried and legal and illegal salaried employment in response to changes in the VAT and CSI + NCSI taxes and subsidies. In this context, we estimate the same VAT reform but maintain CSI + NCSI taxes and subsidies. As expected, in this case informal employment increases in response to the higher VAT (as firms evade more and self-employment and family firms expand), so the revenue to GDP ratio increases only from 0.038 to 0.068, as opposed to the increase to 0.071 when CSI + NCSI taxes and subsidies are eliminated. This result highlights the importance of the simultaneous reduction in CSI + NCSI taxes and subsidies to reduce informality in the face of the VAT reform. But, more importantly for our purposes, it provides reassurance to policymakers that even if, after eliminating CSI taxes and subsidies, labor is re-allocated in the opposite direction as that predicted by our model—an overly pessimistic result contrary to theory and evidence—, the associated loss in VAT revenues is in the order of 0.3% of GDP.

Our proposal has implications for productivity. A growing literature focuses on policies that by inducing resource misallocations generate productivity losses (Gollin, 2006; Restuccia and Rogerson, 2008; Guner et al., 2008; Hsieh and Klenow, 2009a, 2009b; Leal, 2010). In our case, misallocations arise from differences in VAT rates between sectors, differences in the price of labor depending on whether the contractual relation is salaried or non-salaried, and differences in the expected price of salaried labor across firms of different size given that the expected costs of evasion increase with firm size. The proposal eliminates these differences, and results in better factor allocation and, in principle, productivity gains. We do not quantify

---

Throughout the paper we assume an inelastic labor supply. There is a large literature on the impact of taxes on participation rates (e.g., OECD (2011)); on the other hand, the impact on the formal/informal and salaried/non-salaried composition of the labor force is relatively less studied and, for Mexico, central.
these gains, but note only that our proposal would result in higher productivity. This is also a critical result because, first, lagging productivity growth is the main reason why Mexico’s growth performance has been lackluster over the last decades (Levy, 2010); and second, because in Mexico the distortions implicit in the formal-informal dichotomy are associated with large total factor productivity losses (Busso et al., 2011).

Bismarckian SI is the norm rather than the exception in Latin America. Many countries in the Region are also characterized by large informality and high tax evasion. In parallel, in some of these countries NCSI programs are also present, which one can reasonably expect to expand in the years ahead as the Region’s governments respond to the desire for increased coverage of SI and, more generally, greater social equity. Under the present CSI + NCSI architecture, many of these countries may be expanding the coverage of SI at the cost of higher fiscal vulnerability and larger productivity-reducing distortions. Thus, while the focus of the paper is on Mexico, we believe the issues discussed here have broader relevance.

The rest of the paper is organized as follows. Section 2 contains a brief discussion of Mexican SI policies and informality. Section 3 presents relevant stylized facts on tax evasion, informality and labor mobility. Section 4 builds a simple model to capture the behavior described above, while data and calibration are discussed in section 5. Section 6 studies the distortionary effects of CSI + NCSI taxes and subsidies. Sections 7 and 8 present our proposal: section 7 describes the impact of a fiscal reform to finance USI on the government’s fiscal balance and worker’s welfare; section 8 addresses coverage of risks that are specific to salaried workers and complementary pension benefits. Section 9 briefly touches on the implications of our proposal for poverty and productivity. Section 10 discusses implementation issues. Section 11 presents concluding thoughts.

2. Informality, illegality and social insurance

“Informality is a term that has the dubious distinction of combining maximum policy importance and political salience with minimal conceptual clarity and coherence in the analytical literature” (Kanbur, 2009). In this context, we follow Kanbur and define informality with respect to the inobservance of a particular regulation. Given our focus on SI, the relevant regulation is coverage of CSI.

Our definition of informality allows identifying clearly formal and informal workers, since at any point in time an individual worker is either enrolled in CSI or not. The same does
not hold for firms, however, since at the same point of time an individual firm may enroll only a subset of its workers with CSI; in this case, therefore, it is better to think of a formality-informality continuum, with individual firms located in this continuum as a function of the degree of evasion of CSI taxes. The impossibility of defining firms as unambiguously formal or informal implies in turn that the formal and informal sectors cannot be measured with precision. In some general sense, however, it is clear that as the number of informal workers increases (salaried or non-salaried), and firm evasion of CSI taxes grows, the size of the informal sector expands.

Informality and illegality are not equivalent. Even if VA and CSI taxes were perfectly enforced, there would be self-employment and family firms, and thus an informal sector as some workers would not be covered by CSI; differently put, there would be no illegality but there would be informality. On the other hand, illegality can occur with respect to either CSI or VA taxes. Because for the reasons explained below firm’s evasion of both taxes is correlated, informality and illegality will overlap. That said, firms may comply with VAT but evade CSI taxes; this is relevant in a context like Mexico’s, where firms face different VAT rates depending on the good they produce and, more particularly, because some goods are exempt from VAT. In this case there is legality with regards to the VAT and illegality with regards to CSI taxes.

Informality is not defined by the size of firms. This is because depending on the probabilities of detection, some firms may enroll all of their workers in CSI while others may not. Of course, to the extent that these probabilities are correlated with the size of the firm, the tendency will be to find proportionately more small firms in the informal end of the formal-informal continuum, and proportionately more large firms on the other end. But note that the distribution of firms in this continuum is endogenous to the incentives they face; differently put, the same firm may be more or less formal depending on the configuration of CSI, NCSI and VA taxes and subsidies, given enforcement.

Informality is also not equal to self-employment and or employment in firms, precisely because some salaried workers may not be covered by CSI. Nor is informality, finally, equivalent to poverty because, on one hand, the evidence shows that some poor salaried workers are enrolled in CSI; and because many non-poor workers are either non-salaried or not registered by the firms that hire them with CSI, on the other. In fact, there are almost
twice as many informal workers in Mexico as there are poor workers (although most poor workers are informal; see Levy, 2008).

The distinction between informality and illegality is important from the perspective of SI because in Mexico CSI and NCSI programs are not the same. Workers enrolled in CSI are obligated to purchase a bundle of benefits that cannot be separated; workers covered by NCSI can voluntarily access any of the benefits that are freely offered. As a result, only in the first case is the government assured that workers are protected against the risks that it considers relevant. Indeed, if the scope and obligatory nature of CSI and NCSI programs were the same, the distinction between formality and informality would not be relevant; workers would always be covered against the same risks regardless of their salaried or non-salaried status, and regardless of firm’s behavior. On the other hand, even if this was the case, it would still be relevant if firms evaded their tax obligations. Differently put, informality is a concern from the perspective of the government’s social objectives; illegality from the perspective of its fiscal objectives.

3. Four relevant stylized facts on Mexico

3.1 Low Tax Revenues and High VAT Evasion

Mexico’s tax system is strongly centralized, as approximately 96% of total revenues are levied and collected by the federal government (including income, value added, foreign trade and most excise taxes). In addition, the tax code is characterized by many exemptions and special regimes which, combined with imperfect enforcement, translates into a low tax to GDP ratio (Antón and Hernández, 2010; Elizondo-Mayer, 2010). In 2008 this ratio equaled 0.098 compared to a Latin American average of 0.173.

Special treatments are particularly relevant with respect to VAT. The general rate is 16 per cent (15 percent in 2008), but food and medicines are zero-taxed while other services are tax-exempted (education, cultural activities, private medical expenses, some financial services, books and magazines). Furthermore, rates at Mexico’s borders zones are 11 per cent (10 in 2008). These features, combined with high levels of evasion, resulted in a VAT to GDP ratio of 0.037 in 2008. Importantly, Antón and Hernández (2010) estimate that with the current

---

10 Thus, only under CSI are workers forced to save for their retirement pension and forced to purchase life and disability insurance; see Levy (2008) for a fuller discussion.
exemptions but without evasion this ratio would equal 0.061, suggesting that in 2008 VAT evasion resulted in losses of revenue of 2.4% of GDP.

Special tax regimes go beyond sectors. Self-employed workers and family firms may be exempt from the VAT and income tax regime. Instead, these two taxes can be substituted by one on gross sales as long as sales are below an exogenous threshold; a regime known as Repeco, the Spanish acronym for Régimen de Pequeños Contribuyentes (roughly, Minor Taxpayer Regime). Three features of this regime are relevant. First, it prohibits firms from issuing invoices to firms who pay taxes under the normal VAT regime. Second, firms whose sales increase beyond the exogenous threshold and thus pay taxes under the normal VA and income tax regime cannot re-enter as Repecos even if their sales contract and fall below the same threshold. These two features serve to, de facto, permanently segment firms between those that contribute as Repecos and those that contribute under the normal regime. The third feature is evasion. While there are no estimates of the potential revenues that under full compliance would be collected from Repecos, it is nonetheless surprising that in 2008 total collections were only 0.0004% of GDP.

3.2 High firm informality and skewed size distribution

Workers in Mexico can only be enrolled in CSI if the firm that hires them is registered with the Mexican Institute of Social Security (henceforth IMSS, for its Spanish acronym). As a result, the number of firms registered with IMSS can be used as a proxy of firm formality, even if firms do not fully pay CSI taxes. On the other hand, the Economic Census captures firms in Mexico regardless of whether they are registered with IMSS or not. That said, we note that despite its name, the Census only captures economic activity taking place in fixed establishments in urban areas; urban employment in the streets and other non-fixed locations is excluded, as is rural employment. Table 1 compares the number of establishments captured in the 2008 Census with those registered with IMSS in the same year. According to the Census,

---

1 Valero and Sánchez (2010) show that the co-existence of the Repeco and normal VA and income tax regimes introduces a discontinuity in the profit function of tax compliant firms, implying that if both regimes were fully observed there is a range of sales volumes for which profit maximizing firms would not exist. Note also that the prohibition for firms under the Repeco regime to issue invoices to firms under the normal VAT regime implies the breakdown of the value added chain for purposes of accreditation of the VAT. In particular, tax-compliant firms under the normal VAT regime will be dissuaded from purchasing intermediate inputs from firms under the Repeco regime.
in 2008 there were 3,724,019 establishments; IMSS, however, only recorded 795,466 establishments in that year.\footnote{12}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Firm size} & \textbf{Census} & \textbf{IMSS} & \textbf{IMSS/Census} & \textbf{Share in Census Total} \\
(\# of workers) & & & & \\
\hline
1 – 5 & 3,312,092 & 542,064 & 0.16 & 0.89 \\
6 – 10 & 224,086 & 101,231 & 0.45 & 0.06 \\
11 – 50 & 149,968 & 113,458 & 0.76 & 0.04 \\
51+ & 37,873 & 38,713 & 1.02 & 0.01 \\
\hline
Total & 3,724,019 & 795,466 & 0.21 & 1.00 \\
\hline
\end{tabular}
\caption{Registries of Firms, 2008}
\end{table}

As can be seen, the discrepancy between IMSS and Census data is inversely associated with firm size (measured by the number of workers); firm compliance with IMSS registration increases with size. On the other hand, note that 89\% of firms captured by the Census have less than five workers, 95\% less than ten, and only 1\% more than fifty. The picture that emerges from table 1 is that of an economy characterized by a skewed size distribution of firms with, on one end, a large number of mostly small informal firms evading or eluding CSI taxes (and by-and-large paying no VAT with perhaps a few contributing under the Repeco regime); on the other end, a very small number of mostly formal large establishments complying with their CSI and VA and income taxes; and, in the middle, semi-formal firms partly complying with both CSI and VA and income taxes.\footnote{13}

3.3 \textit{High Informal Employment}

Table 2 presents the distribution of employment by firm size and formality status in 2008. Panel A corresponds to urban employment in firms included in the Economic Census.

\footnotetext{12}{The correspondence between Census and IMSS data is imperfect. The unit of observation in the Census is the establishment; for IMSS it is the firm. Some firms may have more than one establishment, so the comparison may overestimate the number of unregistered firms. This problem affects mostly very large firms, as most small firms have only one establishment. On the other hand, some firms may have more than one registry in IMSS. Again, this occurs mostly for very large firms, who for accounting or tax purposes report as two or more legally separate firms (this explains why there are 2\% more firms with 50 or more workers in the IMSS than in the Census data).}

\footnotetext{13}{Leal (2010) compares the distribution of employment by firm size in Mexico with the United States. When only employment in formal firms is compared, the two distributions overlap considerably; when employment in informal firms in Mexico is added, however, the distributions differ importantly as a result of large left tail of employment in mostly small and informal firms.}
Panels B and C contain, respectively, urban employment not associated with firms included in the Census and rural employment (see Appendix 2 for details of data and methodology).\footnote{The occupied labor force includes also 4.6 million public sector workers who, however, have their separate SI regime. In the reminder of the paper we focus only in non-public employment.}

### Table 2: Total private occupied labor force, 2008*

<table>
<thead>
<tr>
<th>Firm Size</th>
<th>Formal</th>
<th>Informal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Urban employment captured in Census</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – 5</td>
<td>596</td>
<td>8,174</td>
<td>8,770</td>
</tr>
<tr>
<td>6 – 10</td>
<td>733</td>
<td>981</td>
<td>1,714</td>
</tr>
<tr>
<td>11 – 50</td>
<td>2,731</td>
<td>1,060</td>
<td>3,791</td>
</tr>
<tr>
<td>50+</td>
<td>4,665</td>
<td>687</td>
<td>5,352</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8,725</strong></td>
<td><strong>10,902</strong></td>
<td><strong>19,629</strong></td>
</tr>
<tr>
<td><strong>Panel B: Urban employment not captured in Census</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-employment</td>
<td>9</td>
<td>4,064</td>
<td>4,073</td>
</tr>
<tr>
<td>2 – 5</td>
<td>213</td>
<td>6,015</td>
<td>6,228</td>
</tr>
<tr>
<td>6 +</td>
<td>1,517</td>
<td>1,403</td>
<td>2,920</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,739</strong></td>
<td><strong>11,482</strong></td>
<td><strong>13,223</strong></td>
</tr>
<tr>
<td><strong>Panel C: Rural employment not captured in Census</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10,747</strong></td>
<td><strong>27,738</strong></td>
<td><strong>38,485</strong></td>
</tr>
</tbody>
</table>

*Thousands of workers; **Distribution by size not available.

Various aspects are of interest. First, the Census underestimates economic activity in Mexico; in fact, it only accounts for 51% of all private workers (19.6/38.4 million). Second, even within employment captured by the Census, informal employment is larger than formal (10.9 vs. 8.7 million); in addition, note that informality is inversely related to firm size, a fact that is consistent with firm evasion of CSI taxes noted in table 1. Third, the majority of workers not captured by the Census are informal (16.8 out of 18.8 million in panels B and C). All in all, 72% of private employment is informal, with a large share accounted by self-employed workers, or by firms with up to five workers (47% of all private employment). Thus, table 2 complements table 1 and indicates that informal employment in Mexico occurs both in firms (presumably illegal) or in establishments without a fixed location (street markets and stands), or by self-employed individuals (street vendors and the like, and rural workers).
3.4 High transits between formal and informal employment

Lastly, Mexico’s labor market is characterized by large mobility of individual workers between formal and informal status. Table 3 exploits the panel structure of Mexico’s employment survey to separate job and status changes between 2007 and 2008. The data is presented as a transition matrix, with the rows referring to the status of workers in 2007, and the columns to the status of the same workers one year later.

### Table 3: Workers job and status change, 2007-2008*

<table>
<thead>
<tr>
<th>Status in 2007</th>
<th>Status in 2008</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No job or status change</td>
<td>Status change, no job change</td>
</tr>
<tr>
<td>Formal</td>
<td>80.1</td>
<td>8.6</td>
</tr>
<tr>
<td>Informal</td>
<td>77.3</td>
<td>6.0</td>
</tr>
</tbody>
</table>

*shares.

As can be seen, 80% of workers who were formal in 2007 kept the same job and their formal status one year later, while 20% changed either job or status or both. Of the latter, 8.6% had the same job one year later, but transited to informal status. This change, which at first sight may sound contradictory, is a reflection of firm’s evasion strategies, as they only register a subset of their workers with CSI and rotate who they register. An additional 7.1% changed jobs, but kept their formal status. This change is what is normally considered as churning, as a worker changes jobs from one formal firm to another, but with no implications for the coverage of SI. Finally, 4.2% changed both jobs and formality status, meaning that they changed from a job with a firm that registered them with CSI to another job with a firm that failed to register them with CSI, or to self-employment or a family firm. Similarly, 77% of workers who were informal in 2007 kept that status one year later; 6% kept the same job, but transited to formality (presumably because their firm registered them with CSI); 13% changed into another informal job; and 3% into a formal job.

On average, 11% of all workers changed status in one year, with or without a job change. This phenomenon, which needs to be distinguished from the normal churning of a labor market, implies that 11% of workers transited in one year between coverage of CSI and coverage of NCSI. More generally, and over longer time spans, Mexican workers experience

---

15 Levy (2008) reports that IMSS experiences about 15 million changes in worker registration a year, with a total stock of formal workers of around 13 million. Some of these registration changes reflect true hiring and firing as firms adjust employment to various shocks, but some also reflect registration rotation without any changes in the individuals working in the firm.
various episodes of CSI and NCSI coverage during their working life, with the associated implications for the coverage of risks at the individual level, risk pooling at the aggregate level, and the accumulation of savings for retirement at both individual and aggregate levels.\textsuperscript{16}

4. The model

We model an economy with special tax regimes, imperfect enforcement and dualistic SI. Here we only describe the features needed to interpret our empirical results; Appendix 1 gives further details and derivations. The economy produces two intermediate and two final goods. Intermediate goods I\textsubscript{1} and I\textsubscript{2} are produced with capital and labor by firms and workers in salaried contractual relations; workers receive CSI benefits when firms comply with the Law, and NCSI benefits otherwise. I\textsubscript{1} and I\textsubscript{2} are aggregated into a final consumption good A, which for simplicity is only produced with intermediate inputs. The second consumption good, B, is produced without intermediate inputs by self-employed workers, or by workers in family firms with non-salaried contractual relations, receiving NCSI. Value added from A, B and intermediate goods I\textsubscript{1} and I\textsubscript{2} makes up total GDP. We focus attention on the labor market and assume the economy is small and open to world markets, so producer prices p\textsubscript{1}, p\textsubscript{2} and p\textsubscript{B} are given by the world market.\textsuperscript{17}

Firms producing I\textsubscript{1}, I\textsubscript{2} and A are legally required to pay income and VA taxes under the regular regime, though they may partly evade these obligations.\textsuperscript{18} VAT rates for I\textsubscript{1} and I\textsubscript{2} differ: I\textsubscript{1} is assumed to be exempt (as is currently the case with food, medicines and the like); I\textsubscript{2} (all other goods) pays VAT at the rate of 15%. Firms producing I\textsubscript{1} and I\textsubscript{2} pay CSI taxes when they hire workers legally (and implicitly receive a subsidy from NCSI when they hire workers

\textsuperscript{16} Levy (2008) uses the IMSS records to construct a panel data of all workers registered in IMSS in 1997 and follows them over the 1997-2006 period, dividing workers between those earning high and low wages (more or less than three minimum wages). He finds that high wage workers spent 77\% of that decade in formality, and low wage workers 49\%. He also finds greater frequency of entry and exit from formality for low than for high wage workers. These results are consistent with data of contribution densities to individual retirement accounts, which in the same period averaged about 45\%, and which were lower for low wage workers.

\textsuperscript{17} Prices are normalized such that p\textsubscript{1} + p\textsubscript{2} + p\textsubscript{B} = 1, and the model satisfies the usual condition of homogeneity of degree zero (one) of all real (nominal) variables; the price of A follows from p\textsubscript{1} and p\textsubscript{2}; see equation (20) below. One could also treat good B as non-traded. This considerably complicates the model without much additional insight. As discussed in Appendix 1, a negatively-sloped demand curve for B would partly offset the labor re-allocations produced by changes in taxes and subsidies as p\textsubscript{B} responds to changes in employment in this sector.

\textsuperscript{18} Income taxes are included here to capture the fact that as the tax base changes revenues from corporate income taxes change as well. But the corporate income tax does not affect behavior, and we ignore personal income taxes.
illegally; see below). Workers producing good B are non-salaried and thus pay no CSI taxes (and receive the same implicit subsidy from NCSI as illegally hired salaried workers). Production of B is assumed to be subject to the Repeco regime. However, since this regime is practically unenforceable, we assume for simplicity that sector B faces no tax.\footnote{Enforcement of the Repeco is left to state governments, as opposed to enforcement of the regular income and VAT regime for the rest of the economy (in our model, firms producing I$_1$, I$_2$ and A), which falls on the Federal Government. States have few incentives to collect given high enforcement costs, non-credible sanction mechanisms and the fact that revenue-sharing formulas provide them with large transfers from the Federal Government. Thus, to simplify we treat sector B as non-taxed. We would get basically the same results if we treated it as taxed but with lower probabilities of detection than in sectors (I$_1$, I$_2$, A), but with little additional insight.} In what follows we refer to sectors (I$_1$, I$_2$, A) as the taxed sector of the economy (although evasion can occur), and sector B as the non-taxed. Table 4 summarizes this information and introduces notation on taxes and subsidies.

**Table 4: Taxes and subsidies on goods and factors**

<table>
<thead>
<tr>
<th>Goods</th>
<th>Income tax</th>
<th>VAT</th>
<th>Labor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Formal</td>
</tr>
<tr>
<td>I$_1$</td>
<td>$\tau^IT$</td>
<td>$\tau_1^{VAT}$</td>
<td>$\tau^{CSI}$</td>
</tr>
<tr>
<td>I$_2$</td>
<td>$\tau^IT$</td>
<td>$\tau_2^{VAT}$</td>
<td>$\tau^{CSI}$</td>
</tr>
<tr>
<td>A</td>
<td>$\tau^IT$</td>
<td>$\tau_A^{VAT} = \gamma \tau_1^{VAT} + (1-\gamma) \tau_2^{VAT}$</td>
<td>n.a.</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: $\gamma$ is the share of VAT exempt goods in good A; n.a. = not applicable; all taxes and subsidies expressed as rates, except for $\tau^{NCSI}$ which is expressed per worker (see below).

4.1 **The intermediate goods sector**

Intermediate goods, indexed by $z = 1, 2$, are produced by a large number of firms in each sector. Firms behave in a competitive fashion selling their output to sector A at the price $p_z$, and produce it with constant returns to scale technology:

(1) \[ I_z = I_z(K_z, L_z) \]

Capital per firm $K_z$ is given, so the representative firm makes positive profits in equilibrium.\footnote{Alternatively, $K_z$ may be interpreted as entrepreneurial ability as in Lucas (1978).} $K_z$ is continuous and distributed among firms according to a distribution function $F(K_z)$ with

\[ F(K_z) \]
support \( k_z = [K_z, \bar{K}_z] \). The corresponding density is denoted by \( f(k_z) \). Capital endowment in the economy is given by \( \bar{K} \). Since good B is only produced with labor, the economy’s capital resource constraint is:

\[
(2) \quad \int_{K_1} \bar{K}_1 f(K_1) dK_1 + \int_{K_2} \bar{K}_2 f(K_2) dK_2 = \bar{K}
\]

Firms make two critical decisions: how many salaried workers to hire, and the optimal combination of legal and illegal workers. The firm’s total labor input, \( L_z \), is:

\[
(3) \quad L_z = L_{fk} + L_{iz}
\]

where \( L_{fk} \) and \( L_{iz} \) refer to legally and illegally hired workers, respectively; which in this case correspond with formal and informal workers. If firms hire workers formally, they pay the formal wage, \( w_f \), and CSI taxes (expressed as a proportion of the wage). Note, however, that in Mexico a share \( \theta \in [0,1] \) of CSI taxes is paid by the government. Thus:

\[
(4) \text{Unit cost of legally hired salaried workers } = w_f [1 + (1-\theta)\tau^{CSI}].
\]

On the other hand, informal labor occurs precisely because CSI taxes are evaded, so labor costs are simply given by the informal wage rate \( w_i \). However, if a firm evades CSI taxes, there is an endogenous probability \( \lambda_z^{CSI}(L_z, K_z) \) that it will be discovered by the authority (Levy, 2008). This probability is proportional to firm’s size as measured by both the amount of informal labor employed and the firm’s capital. For example, if relatively little labor is required to conduct business, the probability that such firm is discovered evading CSI taxes is near zero, and the firm will hire mostly informal workers. In contrast, if conducting business requires hiring many workers, the firm will have an incentive to hire mostly formal workers as the probability of being discovered evading by the authority is high. The probability of detection \( \lambda_z^{CSI}(L_z, K_z) \) is assumed increasing in both arguments. In general, firms will demand both formal and informal workers, but larger firms will demand relatively more formal workers.

---

21 This reflects the “tripartite” nature of CSI (with contributions from firms, workers and the government). In particular, the government subsidizes the health, life and disability insurance and retirement pensions components of CSI. All in all, 16% of CSI taxes are paid by the government; see Appendix 2.
Following Mexico’s Social Security Law, firms caught cheating must pay CSI taxes due on illegally hired workers, times a penalty \( \sigma^{CSI} > 1 \) to dissuade them from evading. Thus:

\[
(5) \quad \text{Expected unit costs of illegally hired salaried workers} = w_i + \lambda^{CSI} \cdot \phi
\]

where \( \phi = [\sigma^{CSI} (1 - \theta) \tau^{CSI} w_i] \). Importantly, this cost increases with the number of illegally hired workers.

Firms must also pay income and value-added taxes. As with CSI taxes, firms have an incentive to evade and face a probability \( \lambda^{VAT} (K_z) \) of being detected by the authority, which is assumed to be an increasing function of the physical capital \( K_z \) of the firm.\(^2\) In turn, if a firm is detected it has to cover the amount of taxes evaded times a penalty \( \sigma^{VAT} > 1 \). Thus the expected VAT payment for a firm of size \( K_z \) is \( \sigma^{VAT} \tau^{VAT} \lambda^{VAT} (K_z) VA(K_z) \), where \( VA(K_z) \) denotes value added. Similarly, the corresponding expected income tax payment is \( \sigma^{IT} \tau^{IT} \lambda^{IT} (K_z) \Pi(K_z) \), where \( \Pi(K_z) \) is the gross profit for a firm of capital size \( K_z \).

We define the rate of VAT compliance, \( \xi^{VAT} (K_z) \in [0,1] \), as the ratio of a firm’s expected VAT payment over the VAT that would be paid under full observance of the Law:

\[
(6) \quad \xi^{VAT} (K_z) = \frac{\min \{ \sigma^{VAT} \tau^{VAT} \lambda^{VAT} (K_z) VA(K_z) \}}{\tau^{VAT} VA(K_z)}
\]

Calculating a similar ratio for income taxes, we find that for a firm of size \( K_z \) the effective tax rates are \( \tau^{VAT} \xi^{VAT} (K_z) \) and \( \tau^{IT} \xi^{IT} (K_z) \).

We are now ready to define the problem of a representative firm in the intermediate goods sector. For given capital \( K_z \), a firm must choose the amount of formal and informal labor, \( \{L_{fz}, L_{iz} \} \), to maximize expected profits:

\[
(7) \quad \text{Max} \Pi_z = \left[ \max \left\{ \frac{1 - \tau^{IT} \xi^{IT} (K_z)}{\xi^{IT} (K_z)} \right\} \right] \cdot \left\{ \left[ \max \left\{ \frac{1 - \tau^{VAT} \xi^{VAT} (K_z)}{\xi^{VAT} (K_z)} \right\} \right] p_z \cdot L_z \cdot \left[ K_z, (L_{fz} + L_{iz}) \right] \right.
\]

\[
- \left[ 1 + (1 - \theta) \tau^{CSI} w_i \right] \cdot L_{fz} - \left[ w_i + \phi \lambda^{CSI} w_i \right] L_{iz}
\]

\(^2\) This assumption reflects the fact that in Mexico collections of CSI taxes and value-added/income taxes are performed by different government agencies. However, the fact that both probabilities of detection depend on the amount of capital implies a correlation between them.
The solution is a pair of demand functions for formal and informal labor of the form:

\[ L_f = L_f[z, p^VAT, \tau^CSI, \theta, w_f(\tau^CSI, \tau^NCSI), w_i(\tau^CSI, \tau^NCSI)] \]

\[ L_z = L_z[z, p^VAT, \tau^CSI, \theta, w_f(\tau^CSI, \tau^NCSI), w_i(\tau^CSI, \tau^NCSI)] \]

Equations (8) and (9) are very important. They capture how firms combine formal and informal (illegal) salaried labor in a context of evasion. As expected, demand for both types of labor increases with output price and decreases with the VAT rate, as the net price received by firms falls (for a given level of evasion). However, given output price and VAT rates and the government’s enforcement efforts, the composition of the firm’s labor demand depends on CSI and NCSI taxes and subsidies, as firms optimally mix formal and informal workers. Higher CSI taxes increase the price of formal vs. informal labor and induce firms to hire more informal workers; conversely, higher subsidies for CSI, \( \theta \), lower the relative price of formal vs. informal labor and induce more formal hiring. On the other hand, individual firms take wage rates for formal and informal labor as given. However, as elaborated below, these wage rates depend on the level of CSI and NCSI benefits, and on how workers value those benefits. Thus, even though NCSI benefits do not appear directly in the firm’s labor demand functions (8) and (9), they do so indirectly through their effect on wages; as a result, the level and composition of firms’ labor demands depends on the dual structure of SI.

We next define the aggregate rate of tax compliance in the intermediate goods sector, \( \bar{\xi}_z^{VAT} \in [0,1] \), as total revenue collected over potential revenue (if all firms fully complied with their tax obligations):

\[ \bar{\xi}_z^{VAT} = \frac{\tau^VAT \int_{z}^{K} \xi_z^{VAT}(K_z)VA(K_z)f(K_z)dK_z}{\tau^VAT \int_{z}^{K} VA(K_z)f(K_z)dK_z} \]

with a similar expression applying to income taxes.

The aggregate compliance rate determines the extent to which the VAT impacts intermediate goods prices. In particular, the “VAT included” prices of \( I_1 \) and \( I_2 \) are:
reflecting the fact that under imperfect compliance the VAT is not fully transmitted to prices.

4.2 The final good sector

This sector is composed of a large number of representative firms that behave in a competitive fashion. Firms use intermediate goods $I_1$ and $I_2$ to produce $A$, and must pay income and value-added taxes. Firms can also evade and, as before, $\xi_A^{VAT}$ and $\xi_A^{IT}$ are the rates of compliance with each tax, which are assumed to be a weighted average of the corresponding aggregate rates of compliance in the intermediate goods sector. Thus, the tax rates effectively paid by firms in the final sector are $\tau_A^{VAT} \xi_A^{VAT}$ and $\tau_A^{IT} \xi_A^{IT}$.

VAT in Mexico is collected by the credit method: the tax applies to each sale, and firms in the final good sector receive a credit for the VAT paid in the previous stage of production. Hence if the cost of the intermediate good (before taxes) is $p_z I_z$, the firm in the final good sector receives a tax credit of $\tau_z^{VAT} \xi_z^{VAT} p_z I_z$. As a result, tax evasion by firms in the intermediate sector implies a trade-off for firms in the final sector. On one hand, a lower rate of compliance (i.e., higher evasion) in the intermediate sector implies that taxes effectively paid by firms in the final sector are lower. On the other, lower compliance rates in the intermediate sector translate into lower tax credit claims by final sector firms. In the extreme case where tax evasion in the intermediate sector is zero, the rate of compliance in the final sector is one and, in principle, these firms have the right to a full tax claim (unless final firms understate sales to reduce VAT payments in the last phase of the VA chain). Thus, the intermediate-final good structure of our model gives place to a transmission mechanism of tax evasion between sectors, as in de Paula and Scheinkman (2010). Since tax credits cannot be generated from informal suppliers and tax payments from formal suppliers cannot be used by informal buyers, there is an incentive for informal firms to conduct business with other informal firms. This scheme thus predicts that tax evasion of a firm in the final good sector is correlated to the tax evasion of firms from which it buys intermediate goods.23

---

23 De Paula and Scheinkman (2010) present empirical evidence for Brazil supporting this idea, labeled by them as “chain effects”. In table 13 below we illustrate that these effects are quantitatively relevant for Mexico.
In such context, the problem of a representative firm in the final good sector is to choose intermediate goods \(\{I_1, I_2\}\) to maximize expected profits:

\[
\text{Max} \prod_A = (1 - \tau^{IT} \xi_A^{IT}) \left\{ (1 - \tau_A^{VAT} \xi_A^{VAT}) p_A A(I_1, I_2) - \sum_{z=1}^{2} (1 - \tau_z^{VAT} \xi_z^{VAT}) p_z I_z \right\}
\]

taking prices \(\{p_1, p_2\}\), tax rates \(\{\tau^{IT}, \tau_1^{VAT}, \tau_2^{VAT}\}\) and rates of compliance by firms in the intermediate sector \(\{\xi_z^{VAT}, \xi_z^{IT}\}_{z=1,2}\) as given.\(^{24}\) The solution is a pair of demand functions for \(I_1\) and \(I_2\), whose proportions will depend on relative prices and the substitution possibilities allowed by the technology. Appendix 1 shows that when this function takes the standard CES form, these proportions are:

\[
\frac{I_1}{I_2} = \left[ \frac{\gamma}{1 - \gamma} \left( \frac{1 - \tau_1^{VAT} \xi_1^{VAT}}{1 - \tau_2^{VAT} \xi_2^{VAT}} \right) \left( \frac{p_1}{p_2} \right) \right]^{\mu(1 - \mu)}
\]

where \(\gamma\) and \(\mu\) are parameters of the production function.

This expression indicates that the relative sizes of the two intermediate good sectors are affected by the degree of compliance with the VAT in each sector, as well as by the existence of differentiated VAT rates. In particular, since \(\tau_1^{VAT} = 0\) as a result of the exemption given to food and medicines and related necessities, there is an underlying distortion which negatively affects the size of the non-exempt sector. This is an important observation, as the fiscal reform contemplated in section 7 consists precisely in eliminating this exemption.

4.3 The final good B sector

This sector captures economic activity by self-employed workers and workers in family firms. The key aspect, of course, is that these workers are not obligated to enroll in CSI. The cost of labor is simply the informal wage \(w_i\), and we have:

\[
\text{(14) Unit cost of non-salaried labor} = w_i
\]

\(^{24}\) When evasion rates are zero \((\xi_z^{VAT} = \xi_z^{IT} = 1\) for all \(z\)) and VAT rates identical, the profit function is

\[
\text{Max} \prod_A = (1 - \tau^{IT}) \{ (1 - \tau^{VAT}) p_A A(I_1, I_2) - \sum_{z=1}^{2} p_z I_z \}, \text{ as expected in an economy with no evasion.} \]
Sector B faces the special Repeco tax regime. However, as noted, high enforcement costs and non-credible sanction mechanisms imply that de facto economic activity is not taxed. As a result, the relationship between establishment size and evasion is not central here, and little is lost by treating this sector as consisting of many identical establishments where output is produced only with labor, \( L_B \), which is sold only to final consumers at the exogenous price \( p_B \).\(^{25}\) The production function, \( B = B(L_B) \), is assumed to have decreasing returns to scale. Thus, the profit function is simply \( \max \Pi_B = p_B B(L_B) - w_i L_B \). This leads to a demand for non-salaried labor of the form:

\[(15) \quad L_B = L_B[p_B; w_i(\tau^{CSI}, \tau^{NCSI})] \]

Equation (15) is very important. Notice that, as with the demands for legal and illegal salaried labor in (8) and (9), \( L_B \) depends on CSI and NCSI taxes and subsidies. Of course, self-employed workers and family firms do not directly pay for either \( \tau^{CSI} \) or \( \tau^{NCSI} \); but this fact does not invalidate the more important fact that, as discussed immediately below, wage rates are determined in the labor market, where these taxes and subsidies play a prominent role.

### 4.4 Social insurance, wage rates and the labor market

Equations (8), (9) and (15) determine the demand side of the market for labor. Assuming an inelastic labor supply \( \bar{L} \), equilibrium in the labor market is given by:

\[(16) \quad \int_{\bar{K}_1} \bar{K}_1 f(K_1) dK_1 + \int_{\bar{K}_2} \bar{K}_2 f(K_2) dK_2 + \int_{\bar{K}_1} \bar{K}_1 f(K_1) dK_1 + \int_{\bar{K}_2} \bar{K}_2 f(K_2) dK_2 + L_B = \bar{L} \]

---

\(^{25}\) This is true only as long as sales by self-employed workers or family firms do not exceed the threshold established to qualify for Repeco (else taxation would occur under the normal VAT/IT regime). Our model fails to capture this discontinuity; see Valero and Sánchez (2010) for a proper treatment. The assumption that good B requires no intermediate inputs is not innocuous. If good B required \( I_1 \) and \( I_2 \) as intermediates, a VAT reform like the one we consider below would also serve to indirectly tax sector B, as input prices would increase without firms having the possibility of crediting the higher VAT on inputs since they are in the Repeco regime (or, more likely, are not registered and pay no taxes at all). Thus, the combination of either high evasion and/or the Repeco regime implies that the VAT is not fully a tax on final consumption, and a higher VAT would through this channel serve to tax the informal sector. This important point is emphasized by Keen (2008).
The first two terms are firms' demand for legal salaried labor given by (8); the next two terms are firms' demand for illegal salaried labor given by (9); the fifth term, lastly, is the demand for non-salaried labor given by (15). In turn, formal employment (henceforth denoted as $L_f$) is given by the first two terms, while informal employment by the next three, comprised of illegal salaried employment and legal non-salaried employment (henceforth $L_i$).

When salaried workers are formally employed they are paid the formal wage and receive the benefits of CSI; when they are informally employed (either as illegally hired salaried workers or as non-salaried workers), they are paid the informal wage and receive the benefits of NCSI. As opposed to CSI, the benefits of NCSI are expressed as a monetary value which is simply the amount that the government spends in SI benefits per informal worker. Thus the utility of formal and informal employment are:

\begin{align}
U_f &= w_f (1 + \beta^{CSI} \tau^{CSI}) \\
U_i &= w_i + \beta^{NCSI} \tau^{NCSI}
\end{align}

Parameters $\beta^{CSI}, \beta^{NCSI} \in [0,1]$ capture the value that workers give to SI benefits. They represent the benefit side of SI taxes and subsidies, in the understanding that from the point of view of workers' behavior what matters is how they perceive these benefits. Thus, the fact that CSI benefits are bundled while NCSI benefits are unbundled is important, as in the first case workers are forced to consume a basket of goods and services of fixed composition, while in the second workers are given the option to consume any combination of benefits. Further, matters of quality and access to service are key (particularly for health care), as are workers perceptions of the value of contingent benefits (like death or disability insurance) or benefits that accrue in the long run (like savings for retirement); see Levy (2008) for further discussion.

Table 5 pulls together the information of firms and workers, and contrasts the unit cost of labor on the basis of the labor contracts as given by (4), (5) and (14), on one hand, with the value of those contracts to workers as given by (17) and (18), on the other.

---

26 We could separate the utility of informal employment between that of salaried and non-salaried employment, and assume non-salaried workers also derive utility from not having a boss, having flexible working hours, and so on; see Maloney (1999, 2004). As long as these factors are invariant to wage rates and the benefits of SI, they do not matter for our analysis, and to avoid cluttering the notation we ignore them here.
Table 5: Costs and benefits of labor contracts

<table>
<thead>
<tr>
<th>Labor contract</th>
<th>Firms pay</th>
<th>Workers receive</th>
<th>Implicit tax/subsidy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal salaried (formal)</td>
<td>$w_i + \phi CSI_i$</td>
<td>$w_i + \beta CSI_i$</td>
<td>$1 - (\beta CSI_i + \theta)$</td>
</tr>
<tr>
<td>Illegal salaried (informal)</td>
<td>$w_i + \phi CSI_i$</td>
<td>$w_i + \beta NCSI_i$</td>
<td>$\phi CSI_i - \beta NCSI_i$</td>
</tr>
<tr>
<td>Non-salaried (informal)</td>
<td>$w_i$</td>
<td>$w_i + \beta NCSI_i$</td>
<td>$-\beta NCSI_i$</td>
</tr>
</tbody>
</table>

The last column of table 5 highlights the structure of taxes and subsidies to different labor contracts implicit in the CSI + NCSI configuration. When $(\beta CSI_i + \theta) < 1$ formal labor is subject to a “pure” tax, which arises because workers do not fully value the benefits of CSI on a peso-by-peso basis; as a result, firms pay more for labor than what workers receive (despite the fact that CSI is partly subsidized by the government). On the other hand, in a context in which enforcement of CSI taxes is imperfect, illegal salaried labor is implicitly subsidized by NCSI. Moreover, this subsidy is inversely proportional to firm size, as smaller firms face lower probabilities of detection than larger ones. Finally, non-salaried labor is unambiguously subsidized by NCSI. Note that the implicit subsidy to informal employment is larger for non-salaried vs. salaried labor given that self-employed workers and family firms are not obligated to pay CSI taxes at all, with the difference between the two narrowing as firm size falls (given falling detection costs). It is natural to expect that a tax-cum-subsidy structure that depends on the size of the firms and the type of labor contract will bias the size distribution of firms towards smaller firms, and the distribution of workers towards non-salaried and illegal salaried employment; a bias that is consistent with the size distribution of firms and the composition of employment presented in tables 1 and 2.

A critical question is how CSI + NCSI tax-cum-subsidies impact the labor market. If this market was fully segmented, with no possibility for formal workers to obtain informal jobs, and no possibility for informal workers to obtain formal jobs, these taxes and subsidies would separately impact the formal and the informal wage. For instance, a decrease in CSI taxes would increase firms’ demand for formal workers. Under complete segmentation, informal workers could not become formal because there are barriers to entry into formality; as a result, the formal wage would increase, but the informal wage would not change. Similarly, an increase in NCSI benefits would increase the utility of workers that are self-employed, work
in a family firm or are illegally hired, but firms would not be able to change the composition of their work force in favor of more illegal workers, nor would it induce any salaried workers to enter self-employment or start a family firm, because there are barriers to entry into informal employment; as a result, the formal wage would remain invariant.

The empirical evidence for Mexico does not support the hypothesis of a completely segmented labor market, with no transits of individual workers between formal and informal status; on the contrary, large formal-informal mobility has been extensively documented. A key implication is that formal and informal wage rates cannot be determined by two independent mechanisms. On the contrary, one expects that as $\tau^{CSI}$ increases the demand for formal labor will fall, increasing the supply of labor to the informal sector and putting downward pressure on the informal wage. Conversely, one expects that a higher $\tau^{NCSI}$ will stimulate the supply of labor to the informal sector, putting upward pressure on the formal wage.

In fact, as shown below in section 6, and in accordance with the available empirical evidence for Mexico, CSI taxes are not fully shifted to formal workers, implying in turn that firms adjust employment as a response to higher labor costs, or that $\partial L_f / \partial \tau^{CSI} < 0$; see Marrufo (2010), Heckman and Pages (2004) and Kugler (2011). On the other hand, more recent evidence shows that for individual programs that comprise $\tau^{NCSI}$ it is the case that $\partial L_i / \partial \tau^{NCSI} > 0$. Here we take advantage of these results, and of the observed large mobility

---

27 See Maloney (1999, 2004), Gong, et al. (2004), Navarro and Schrimpf (2004) and Calderon (2000, 2006). These studies exploit the panel structure of Mexico’s employment survey, allowing them to follow individual worker transits between formal and informal status, finding that in one year between 10 to 15% of workers change labor status; see also table 3. Levy (2008) uses the social security registries to follow individual workers over a ten year period and finds large mobility between the formal and the informal sector. Duval and Smith (2010) test the hypothesis of segmentation in Mexico’s labor market and find that when public sector employment is excluded and segmentation is measured as the share of informal salaried workers unable to find a formal salaried job at any wage, less than 27% of informal workers are rationed from formal jobs. Bell (1997) and Maloney and Nunez (2004), on the other hand, provide evidence that the minimum wage is not binding; Duval and Smith (2010) find that 99% of informal workers have a reservation wage higher than the minimum wage.

28 The available evidence refers to individual NCSI programs, as opposed to CSI programs which almost always are legislated and their effects captured under a single statutory rate $\tau^{CSI}$. Juárez (2009) finds that a health program for female workers in Mexico City increases the probability of informal employment for those with low education and decreases informal wages. Gallani and Gertler (2007) find that a pension program for adults over 70 years of age lowers participation in salaried employment for workers close to retirement age. For a health program for informal workers introduced in 2002 known as Seguro Popular, studies by Barros (2009); Gallardo-Garcia (2006) and Campos and Knox (2010) find no impact on informal employment. However, recent evidence points to positive effects on informal employment or a decrease in the share of formal employment, although the magnitude varies depending on the data used; see Azuara and Marinescu (2011), Duval and Smith (2011) and
of workers between formal and informal status, to assume that worker’s utilities are equalized across forms of employment. Given (17) and (18), this implies a relationship between \( w_f \) and \( w_i \) of the form:

\[
(19) \quad w_i = w_f (1 + \beta^{CSI} \tau^{CSI} - \beta^{NCSI} \tau^{NCSI})
\]

Given this relationship, one can use condition (16) in the labor market to find the equilibrium formal and informal wage rates that are consistent with firm’s and worker’s decisions, given VA, income, CSI and NCSI taxes and subsidies, and the government’s enforcement efforts.

It is useful, finally, to introduce an index of the real wage. In Appendix 1 we show that the “VAT included” price of good A is:

\[
(20) \quad p_A = \left[ (\frac{p_A^1}{\gamma})^{1/(\mu-1)} + \left( \frac{p_A^2}{1-\gamma} \right)^{1/(\mu-1)} \right]^{\mu/(\mu-1)}
\]

Since there is no VAT on good B, \( p_B = p_B \). Thus, the consumer price index, CPI, is:

\[
(21) \quad CPI = \delta p_A + (1-\delta) p_B
\]

where \( \delta \) is the share of final good A in total consumer expenditures. We next define indices of the real formal and informal wage rates, respectively, as \( w_f / CPI \), and \( w_i / CPI \). The important point, of course, is that real wages depend on outcomes in the labor market as determined by the incentives faced by firms and workers coming from the VAT and SI taxes and subsidies, but also directly on the VAT regime as it affects consumer prices.

Aterido et al. (2011), Bosch and Campos (2010) find decreases in formal employment and in the share of formal employment, and Bosch and Cobacho (2011) find decreases in formal employment for young workers in small firms. Pérez-Estrada (2011) finds decreases in formal employment and a decrease in informal wages. See Bosch et al. (2012) for a survey. To the best of our knowledge, however, there are no studies of the joint effects of all NCSI programs on \( L_i \) (which conceptually is what one should measure, and for which we present some estimates in section 6).

The assumption of utility equalization simplifies the math, but is not essential. One can posit as well that for some unspecified reason there is a utility differential across sectors such that \( U_f - U_i = \kappa > 0 \). Then a variant of (19) is \( \Delta L^f = \kappa \left[ \frac{U_f}{U_f^o} / \frac{U_i}{U_i^o} \right]^\eta - \kappa \), where the superscript o refers to an initial equilibrium, \( \Delta L^f \) to the flows of workers between the formal and informal sector after any exogenous change, and \( \eta \geq 0 \) the elasticity of labor flows to utility differentials. The key point is not whether utilities are fully equalized or not, but whether, given any differential (including no differential), at the margin there is any change in labor flows as utilities change.
4.5 The government’s budget constraint under the CSI + NCSI configuration

Given our focus on the fiscal implications of the CSI + NCSI duality, we turn to the
government’s budget constraint. Revenues consist of endogenously determined VA, income
and CSI taxes, plus other exogenous revenues denoted by $\overline{R}$ (including oil rents). The
corresponding expressions for VA and income tax revenues are in Appendix 1, and are derived
aggregating over all firms producing intermediates given the compliance rate of each firm, and
calculating as well VA and income taxes on final good A (netting out VAT paid on
intermediates). In turn, expenditures consist of endogenously determined subsidies to CSI and
NCSI programs, and all other expenditures, which we take as exogenous and denote by $\overline{G}$. Importantly, spending in Progresa-Oportunidades and other targeted programs that transfer
income to the poor, as well as all non-insurance related social spending is included in $\overline{G}$. The
government’s fiscal balance, FB, is:

$$FB = [\overline{R} + R^{\text{VAT}}(\cdot) + R^{\text{IT}}(\cdot) + (1 - \theta)\tau^{\text{CSI}} w_f(\cdot)L_f(\cdot)]$$

$$-\{\overline{G} + (1 - \theta)\tau^{\text{CSI}} w_f(\cdot)L_f(\cdot) + \theta\tau^{\text{CSI}} w_f(\cdot)L_f(\cdot) + \tau^{\text{NCSI}} L_4(\cdot)\}$$

where $\cdot$ highlights the variables that are endogenously determined in our model. Note that
the revenue side only includes CSI taxes paid by firms and workers, $(1 - \theta)\tau^{\text{CSI}} w_f(\cdot)L_f(\cdot)$, while the expenditure side includes all CSI spending (which for clarity we separate into two
parts). As a result, CSI exerts a net pressure on the fiscal balance only to the extent that it is
subsidized.

We now discuss the channels through which the VAT and CSI + NCSI taxes and
subsidies impact the FB in a context of informality and evasion. Consider expenditures first.
Regardless of the formal-informal division of employment, the SI of all workers is subsidized
by the government. From the spending point of view the question is whether the per worker

---

30 In 1997 Mexico transitioned from a pay-as-you-go to a defined contribution system of retirement pensions, as a
result of which the pensions of the transition generation are paid directly from government revenues. This is
included in $\overline{G}$ as this spending has no bearing on firms’ labor costs or workers’ utility. Spending in CSI includes
only subsidies for currently active workers, which affect firms’ labor costs and workers’ utility.
CSI subsidy, $\theta \tau^{\text{CSI}} w_f(...)$, is higher or lower than the per worker NCSI subsidy, $\tau^{\text{NCSI}}$. As shown in Appendix 2, for Mexico in 2008 we find that:

$$\tau^{\text{NCSI}} > \theta \tau^{\text{CSI}} w_f(...)$$

Inequality (23) indicates that from the point of view of spending, the formal-informal composition of the labor force is not irrelevant: *ceteris paribus*, higher informal employment deteriorates the FB.

Consider now the revenue side. Note first that higher VAT rates have a contradictory impact on the FB: on one hand, they increase revenues; on the other, they induce more evasion and informality. Firms in the intermediate goods sector will decrease their demand for salaried labor, with the effect stronger for larger firms (as they can evade less); in parallel, the non-taxed sector B will expand. These two effects erode the tax base and lower VA and income tax revenues. In addition, because higher VAT rates increase informal employment, there is also a positive impact on expenditures, given (23). Thus, although the expectation is that the net effect of higher VAT rates is to increase the FB, this impact is partly offset by higher evasion and higher informal employment.

What about CSI taxes? Given $\beta^{\text{CSI}}$ and $\theta$, an increase in $\tau^{\text{CSI}}$ lowers firms’ demand for salaried labor and tilts its composition in the direction of more informal workers as the relative price of legal to illegal salaried workers increases. Larger firms—which comply more with VA and income taxes—will reduce employment relatively more than smaller ones. As the taxable sector of the economy contracts the same will happen to the tax base. Expenditures are also affected as informal employment expands. On the other hand, an increase in the share of CSI taxes that is subsidized by the government, $\theta$, operates in the opposite direction: from the point of view of firms, legal salaried workers are less costly and thus firms expand their demand for labor and tilt its composition in the direction of formality. As the taxable sector expands, value added and income taxes increase. In parallel, expenditures in CSI increase, of course, although this effect is partly compensated by smaller subsidies to NCSI.

Finally, the effects of augmenting $\tau^{\text{NCSI}}$ are as follows. First, informal employment expands both because sector B is more subsidized, and because firms tilt the composition of

---

31 Inequality (23) follows from the observed growth in spending for NCSI programs since the mid 1990s, and one would expect the difference to widen as these programs continue to expand; see Levy (2008).
their labor force towards more illegal employment (with smaller firms benefiting from the higher subsidy proportionately more than larger firms). Revenues fall as sector B expands and evasion increases in sectors I₁ and I₂. Additionally, expenditures increase as informal employment expands (given (23)), so the net effect is a deterioration of the FB.

In the next sections we turn to quantify these effects. We close here with three observations. First, the discussion highlights the importance of considering the spillover effects of VA taxes on informality and, in turn, the spillover effects of CSI and NCSI taxes and subsidies on VA and income tax revenues. Second, it clarifies that the composition of spending in SI between CSI and NCSI matters not only from the social point of view (as it determines which workers are covered against what risks), but also from a fiscal point of view, as it affects tax revenues and total spending in SI. It also indicates that, while the root cause of informality is found in the CSI + NCSI configuration, in a context of evasion and special tax regimes (and a sector of the economy, B, which is basically out of the reach of the tax authorities), VA taxes can also strengthen informality.

Third, the discussion illustrates the dilemmas faced by SI policy in the context of the CSI + NCSI configuration. From the social point of view, setting $\tau_{NCSI} = 0$ is clearly unacceptable, as it leaves non-salaried workers and illegally hired salaried workers without any protection against risks; on the contrary, from a social point of view one would in principle like to set NCSI benefits close to the levels of CSI benefits, as regardless of their salaried or non-salaried status workers face similar risks. At the same time, however, expanding NCSI programs contracts the tax base and increases the level of informality and illegality in the economy (and lowers productivity).

5. Data and model calibration

We fit our model to Mexico for 2008. Appendix 2 gives details of data sources and calibration; here we only provide a general overview.

5.1 Data

All tax and subsidy rates and penalty parameters are set at their legislated values, except for the income tax rate.³² Per worker NCSI subsidy is calculated dividing observed spending in

---
³² The statutory rate for 2008 is 28%, but the Income Tax Law has many deductions and deferrals associated with capital investments that are not captured in our model. In this case we simply fix the rate to reproduce the same level of revenues observed in 2008 (=10.5%).
NCSI programs by observed informal employment. Data from the 2008 Economic Census is matched with IMSS registries to calculate the distribution of salaried employment by firm size, sector ($I_1$ and $I_2$) and formal-informal composition. The same Census is used to obtain the distribution of the capital stock in each sector and calculate $f(K_1)$ and $f(K_2)$; see equation A2.1 and the discussion in Appendix 2. Numbers on formal employment from IMSS registries are complemented with the Employment Survey to calculate $L_y$ and $\bar{L}$ (excluding public sector workers). We use official fiscal accounts for total revenues from VA, income and CSI taxes. Workers’ valuations of CSI and NCSI, $\beta^{CSI}$ and $\beta^{NCSI}$, are taken from the estimates of Levy (2008). Evasion rates are set parametrically to reproduce the same level of revenues given the tax base and statutory rates. Other parameters are chosen from the National Accounts or, when this is not possible, set to replicate the data. Finally, consistency is checked by calculating GDP on the income and value added side, as well as testing for homogeneity of degree one (zero) for nominal (real) variables with respect to $p_1$, $p_2$ and $p_y$.

5.2 Model fitness

We evaluate the fitness of our model in four dimensions: macroeconomic and fiscal accounts; social insurance accounts; distribution of employment; and VAT evasion rates. Table 6 compares the macroeconomic and fiscal accounts generated by the model with those observed in 2008. The exogenous components of expenditures and revenues, $\bar{G}$ and $\bar{R}$ are by construction set equal to those in the data; the remaining entries are endogenously obtained in the model. For clarity, we net out CSI taxes paid by firms and workers from both the revenue and the expenditure side; see equation (22). As may be seen, these accounts are replicated remarkably well. Note that in 2008 subsidies to NCSI were almost two and half times subsidies to CSI (1.25 vs. 0.52% of GDP).\footnote{Our data only captures spending in NCSI programs by the Federal Government, except for health, where we also include sub-national spending. However, some state governments also subsidize non-contributory pensions, day care and housing programs, which we do not measure here as data is fairly incomplete. Total public spending in NCSI programs is in all likelihood higher than 1.25% of GDP.}

\footnote{Our data only captures spending in NCSI programs by the Federal Government, except for health, where we also include sub-national spending. However, some state governments also subsidize non-contributory pensions, day care and housing programs, which we do not measure here as data is fairly incomplete. Total public spending in NCSI programs is in all likelihood higher than 1.25% of GDP.}
Table 6: Macroeconomic and Fiscal Accounts

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Calibrated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pesos* % GDP</td>
<td>Pesos* % GDP</td>
</tr>
<tr>
<td><strong>GDP</strong></td>
<td>12,110.5</td>
<td>12,146.0</td>
</tr>
<tr>
<td><strong>Public Expenditures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsidies to NCSI</td>
<td>151.5 1.25</td>
<td>151.0 1.24</td>
</tr>
<tr>
<td>Subsidies to CSI</td>
<td>63.7 0.52</td>
<td>62.2 0.51</td>
</tr>
<tr>
<td>Other exogenous</td>
<td>2,679.5 22.1</td>
<td>2,679.5 22.0</td>
</tr>
<tr>
<td><strong>Public Revenues</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value-added tax</td>
<td>457.2 3.77</td>
<td>457.9 3.76</td>
</tr>
<tr>
<td>Corporate income tax</td>
<td>393.0 3.24</td>
<td>392.4 3.23</td>
</tr>
<tr>
<td>Other exogenous</td>
<td>1,852.0 15.29</td>
<td>1,852.0 15.24</td>
</tr>
<tr>
<td><strong>Fiscal balance</strong>**</td>
<td>-192.5 -1.58</td>
<td>-190.4 -1.56</td>
</tr>
</tbody>
</table>

*Thousands of millions of 2008 pesos. **Negative is deficit.

Social insurance accounts are presented in Table 7, where we have broken down $\tau^{CSI}$ and $\tau^{NCSI}$ into their various components. For the case of CSI: health, disability and work-risk insurance and day care services (under the IMSS heading, as this is the entity in charge of their administration); retirement pensions (under Afores, the private firms that administer workers retirement savings), housing (Infonavit, the housing institute for formal workers), and state labor taxes (which, while not directly related to SI benefits, are bundled with CSI taxes). For the case of NCSI: health, retirement pensions, housing, and day care (and no state taxes). Again, the model closely replicates the data for all items.

Table 7: Social Insurance Accounts*

<table>
<thead>
<tr>
<th></th>
<th>Firms and workers</th>
<th>Government</th>
<th>Firms and workers</th>
<th>Government</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMSS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>109.3</td>
<td>44.3</td>
<td>111.9</td>
<td>43.2</td>
</tr>
<tr>
<td>Disability</td>
<td>21.0</td>
<td>1.9</td>
<td>21.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Work-risk</td>
<td>16.8</td>
<td>0</td>
<td>17.2</td>
<td>0</td>
</tr>
<tr>
<td>Day care</td>
<td>8.4</td>
<td>0</td>
<td>8.6</td>
<td>0</td>
</tr>
<tr>
<td>Afores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retirement</td>
<td>54.9</td>
<td>17.5</td>
<td>55.4</td>
<td>17.1</td>
</tr>
<tr>
<td>Infonavit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>43.4</td>
<td>0</td>
<td>44.0</td>
<td>0</td>
</tr>
<tr>
<td>State taxes</td>
<td>25.7</td>
<td>0</td>
<td>26.4</td>
<td>0</td>
</tr>
<tr>
<td><strong>B. Non-contributory social insurance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>0</td>
<td>131.0</td>
<td>0</td>
<td>130.6</td>
</tr>
<tr>
<td>Pensions</td>
<td>0</td>
<td>9.5</td>
<td>0</td>
<td>9.5</td>
</tr>
<tr>
<td>Housing</td>
<td>0</td>
<td>9.2</td>
<td>0</td>
<td>9.1</td>
</tr>
<tr>
<td>Day care</td>
<td>0</td>
<td>1.7</td>
<td>0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

*Figures in thousands of million pesos.
Table 8 presents employment data according to different criteria. Panel A divides total employment into the $I_1$ (food, medicines and other necessities exempt from VAT), $I_2$ (all other goods with a VAT rate of 15%) and B (self-employment and family firms) sectors. Panel B according to formality status; panel C by firm size. The model replicates fairly closely the structure of employment, but fails somewhat in the middle of the distribution in the $I_2$ sector. 34

Table 8: Composition of Employment*

<table>
<thead>
<tr>
<th>Panel A: By sector</th>
<th>Observed</th>
<th>Calibrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$ (food, medicines)</td>
<td>5.28</td>
<td>5.24</td>
</tr>
<tr>
<td>$I_2$ (all other)</td>
<td>14.84</td>
<td>14.86</td>
</tr>
<tr>
<td>B (self-employed and family firms)</td>
<td>18.91</td>
<td>18.93</td>
</tr>
<tr>
<td>Total</td>
<td>39.03</td>
<td>39.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: By formality status</th>
<th>Observed</th>
<th>Calibrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td>12.76</td>
<td>12.29</td>
</tr>
<tr>
<td>Informal</td>
<td>26.27</td>
<td>26.73</td>
</tr>
<tr>
<td>Salaried</td>
<td>7.36</td>
<td>7.80</td>
</tr>
<tr>
<td>Non-salaried</td>
<td>18.91</td>
<td>18.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: By firm size</th>
<th>$I_1$ (food, medicines)</th>
<th>$I_2$ (all other goods)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>Calibrated</td>
<td>Observed</td>
</tr>
<tr>
<td>1-5 employees</td>
<td>2.53</td>
<td>2.39</td>
</tr>
<tr>
<td>6-20 employees</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>21-50 employees</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>50-100 employees</td>
<td>0.34</td>
<td>0.40</td>
</tr>
<tr>
<td>101+ employees</td>
<td>1.27</td>
<td>1.29</td>
</tr>
</tbody>
</table>

*Figures in millions of workers. Note: figures do not coincide with those in table 2; see Appendix 2 for details.

Finally, table 9 presents VAT revenues. The calculated revenue to GDP ratio of 0.0376 is very close to the observed value of 0.0377. In parallel, we carry out an exercise where we estimate VAT revenues under full compliance, and compare it with the calculations made from the National Accounts by Antón and Hernández (2010). Finally, we also carry out an exercise where we eliminate exemptions in sector $I_1$, so that both VAT rates are equal at 15% and assume full compliance, and also compare it with the calculations made by Antón and Hernández. The levels of revenues are again very close, indicating that the underlying evasion rates are also close to the actual ones.

34 An important assumption is that all employment in firms captured by the Census is salaried, while all employment in activities excluded from the Census is non-salaried. This is not exactly so. On one hand, there is salaried employment in some rural firms and, probably, in urban firms excluded from the Census; however, we have no data on the number of firms in this situation, nor on their size or capital stock, and as a result treat their workers as non-salaried. On the other, some firms captured in the Census may have non-salaried contractual relations. That said, the aggregate numbers on formal and informal employment are very close to the ones obtained from the Employment Survey; see Appendix 2.
Table 9: VAT Revenues and Evasion
(% of GDP)

<table>
<thead>
<tr>
<th>Observed or calculated from National Accounts</th>
<th>Calibrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1^{VAT} = 0$, $\tau_2^{VAT} = 0.15$</td>
<td></td>
</tr>
<tr>
<td>Under current enforcement</td>
<td>3.77*</td>
</tr>
<tr>
<td>Assuming full compliance</td>
<td>6.08**</td>
</tr>
<tr>
<td></td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td>6.12</td>
</tr>
<tr>
<td>$\tau_1^{VAT} = \tau_2^{VAT} = 0.15$</td>
<td></td>
</tr>
<tr>
<td>Under current enforcement</td>
<td>n.o.</td>
</tr>
<tr>
<td>Assuming full compliance</td>
<td>10.7**</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>10.6</td>
</tr>
</tbody>
</table>

* Observed. ** Calculated from National Accounts by Antón and Hernández (2010); n.o. = not observable.

All in all, the model replicates Mexican data pretty closely along these four dimensions. In the next two sections we proceed, first, to simulate how the economy responds to various parameter changes under the CSI + NCSI configuration and, second, to simulate a fiscal reform that funds replacing this configuration with USI.

6. Four Preliminary Questions

6.1 Who pays for social insurance?

Table 10 decomposes the effects of CSI and NCSI taxes and subsidies on wage rates and worker’s utility (panel A), employment (panel B) and spending and tax revenues (panel C). We take as the benchmark scenario the situation where there is no SI (column two); next, we assume that CSI is introduced with $\tau^{CSI}$ at its current rate of 0.38, but with no subsidies (column three); further, we introduce subsidies to CSI at the current rate ($\theta = 0.16$, implying a subsidy per worker of 5,062 pesos), but assume there is no NCSI (column four); finally, we introduce NCSI at the observed subsidy of 5,650 pesos per worker so this scenario is, de facto, the 2008 status quo (last column). To interpret this and the following tables, we note that the estimated “pure” tax on formal labor is in the order of 20% of the formal wage, while the estimated subsidy to informal labor is in the order of 7% of the informal wage (see table 5).
Table 10: The impact of CSI and NCSI

<table>
<thead>
<tr>
<th>Panel A</th>
<th>No SI</th>
<th>Only CSI</th>
<th>Subsidies to CSI</th>
<th>Status quo: Subsidized CSI + NCSI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wages and utility</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal wage</td>
<td>1.00</td>
<td>0.79</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>Informal wage</td>
<td>1.00</td>
<td>0.88</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>Worker's utility</td>
<td>1.00</td>
<td>0.88</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal salaried</td>
<td>26.04</td>
<td>12.65</td>
<td>13.65</td>
<td>12.29</td>
</tr>
<tr>
<td>Informal salaried</td>
<td>0</td>
<td>7.68</td>
<td>7.97</td>
<td>7.80</td>
</tr>
<tr>
<td>Non-salaried</td>
<td>12.99</td>
<td>18.70</td>
<td>17.41</td>
<td>18.93</td>
</tr>
<tr>
<td><strong>Panel C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fiscal</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsidies to CSI</td>
<td>0</td>
<td>0</td>
<td>67.3</td>
<td>62.2</td>
</tr>
<tr>
<td>Subsidies to NCSI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>151.1</td>
</tr>
<tr>
<td>VAT revenues</td>
<td>496.7</td>
<td>459.5</td>
<td>468.2</td>
<td>457.9</td>
</tr>
<tr>
<td>IT revenues</td>
<td>399.0</td>
<td>392.4</td>
<td>393.8</td>
<td>392.4</td>
</tr>
<tr>
<td>Net fiscal impact#</td>
<td>------</td>
<td>(-) 43.8</td>
<td>(-) 101.0</td>
<td>(-) 258.7</td>
</tr>
</tbody>
</table>

* Index; ** millions of workers; *** thousands of millions of pesos; # vs. the no SI situation.

Start with unsubsidized CSI. First, very importantly, wage rates decline vis-à-vis the no SI case; the formal wage, in particular, by about 21%. An immediate implication is that the incidence of CSI taxes falls mostly on workers: in fact, in this case 70% of CSI taxes are paid by them. This indicates that CSI is basically changing the composition of workers’ consumption (less disposable income and more SI benefits), rather than re-distributing income from firms to them. Differently put, it is as if when workers are formally employed they pay a “VAT ear-marked for SI benefits”, but labeled as “SI Contribution”; a “VAT” of 21% which is additional to a (non ear-marked) VAT that they pay when they make their purchases. At the same time, and usually less noted, CSI taxes also lower informal workers’ wages (by 12%) whom, however, get no SI coverage in return.35

Second, formal workers’ utility falls by less than their wage, 12% vs. 21%. Utility is lower than in the no SI case because workers under value the benefits of CSI. This highlights the critical difference between worker’s individual utility and the government’s welfare function. From the perspective of the government, the CSI equilibrium is presumably better

35 Our findings are consistent with results for other countries in Latin America. IDB (2004, p. 208) notes that “All in all, the available evidence for Latin America suggests that at least part of the costs of the non-wage benefits is passed on to workers in the form of lower wages. A few studies find evidence that workers pay for the entirety of benefits, but the majority find that employers bear a share of the cost...Therefore, the evidence is fairly robust that although a large share of the benefits is likely to be paid by employees, mandatory benefits regulations have a cost in terms of (formal) employment”. See also Heckman and Pages (2004) and Kugler (2011).
than the no SI equilibrium because it is assured that workers consume the bundle of goods that it considers they should consume: purchasing disability, life and health insurance, saving for retirement and for a house, and so on. From workers perspective, on the other hand, the CSI equilibrium is worse than the no SI equilibrium, as their wages are taxed with the proceeds used to give them in return something they do not value as much.

Third, CSI taxes have large revenue-reducing effects. Total revenues from VA and income taxes fall by 5% (or 0.3% of GDP), given increased informality of workers and firms (see below). Thus, while CSI taxes generate revenue, they in parallel induce losses of other sources of revenue by expanding the nontaxable sector of the economy and inducing firms into illegal behavior. Of course, the official fiscal accounts do not register the foregone revenue, as there is no item in the budget labeled “Lost Revenue from Informality”. But the absence of a explicit line-item in the fiscal accounts should not mask that informality is fiscally costly.

Subsidies to CSI partly offset the changes noted above. As the tax on salaried labor is reduced, real wages and worker’s utility increase, and evasion of VA and income taxes falls marginally. On the other hand, the total fiscal cost of CSI is higher, as the (explicitly recorded) cost of subsidies to CSI is added to the foregone (though unrecorded) revenues from VA and income taxes.

NCSI programs, on their part, have the following effects. First, the government is closer to achieving its objective of providing all workers with coverage against risks, as now informal workers receive some SI benefits. Second, because informal workers now get some SI benefits that neither them, nor the firms that may hire them, have to pay for directly, they have higher utility. Third, real wages move in opposite directions: as the supply of labor to the informal sector expands the formal wage increases while the informal one falls. Thus, NCSI programs (like subsidies to CSI) modify the incidence of CSI taxes. In this case, which is the one that corresponds to the 2008 status quo and therefore the relevant one for our purposes, only 64% of CSI taxes are shifted to workers (vs. 70% in the absence of subsidies to CSI and NCSI programs). Fourth, a new item appears and is explicitly recorded in the expenditure side of the budget under the label of subsidies to NCSI programs (or, more precisely, under the label of health, pension, day care programs, and so on, for informal workers); and the

---

36 Our estimate of the net incidence of CSI taxes can be contrasted with Marrufo’s (2010), who analyses the 1997 reform to Mexico’s Social Security Law and finds that 57% of CSI taxes were shifted to workers.
unrecorded item in the revenue side of the budget expands as higher informality increases foregone revenues from VA and income taxes.

We close bringing up a point at times under-emphasized in discussions of the incidence of CSI taxes. Observed wages reflect subsidies to CSI and NCSI worth 1.7% of GDP. As shown, in their absence workers would pay more for CSI benefits. In this context, however, we cannot really establish the full incidence of CSI taxes, because our model is silent about the incidence of the taxes necessary to pay for CSI + NCSI subsidies. Differently put, a richer structure is needed to answer the incidence question. That said, our model does indicate that, contrary to what is often assumed in policy discussions (and more often in political discourse), CSI taxes are not serving to redistribute income from firms to workers.\footnote{It is useful to recall that 89% of all firms have at most five workers. Although we have no precise data, it is likely that many of these are family firms, where redistribution from firm owners to workers would occur within the same household. Even within firms registered with IMSS, the majority has up to five workers; see table 1. The point here, of course, is not that there should be no redistribution from high to low income households, but that CSI taxes are a very coarse redistributive tool, because the assumption that all firm owners belong to high income households and all workers to low income ones is in all likelihood flawed.}

6.2 \textit{What are the effects of SI on employment?}

Table 10 also serves to assess the impact of the CSI + NCSI configuration on the composition of employment (recall that we assume constant participation rates). Begin again with unsubsidized CSI. Because firms do pay for a share of CSI taxes, they reduce their total demand for salaried labor and tilt its composition in the direction of illegal workers. Informal employment expands both because firms now hire salaried workers illegally and because non-salaried employment in sector B grows.\footnote{Although we do not quantify these effects, the change in the composition of employment also increases profits or quasi-rents in the non-salaried sector of the economy (the value of land, street corners), clearly an unintended redistributive impact of CSI.} Relative to the no SI benchmark, salaried employment falls by 22\% and changes in composition, as 36\% of workers are now hired illegally; in parallel, there is a 44\% increase in employment in sector B. These effects are partly reversed when CSI is subsidized. In this case, relative to the no SI equilibrium, salaried employment falls by 17\%, while relative to the unsubsidized CSI case it expands by 6.3\% (while its composition changes slightly in the direction of legality); in parallel, employment in sector B contracts by 7\%
NCSI programs, as expected, lower salaried employment, tilt its composition in the direction of illegality, and promote employment by family firms and self-employed workers. Relative to the case where CSI is subsidized, the joint effect of all NCSI programs is to lower salaried employment by 7% (and increase the share of illegal employment from 37 to 39%); in parallel, employment in sector B grows by 8.7%. As a result, total informal employment grows by 5.3%. We highlight that our model picks up the effects of all NCSI programs simultaneously, in contrast to the papers cited in section 4, which measure the effects of individual programs.39

Contrasting the changes in employment caused by CSI vs. NCSI programs, it follows that the distortion created by the former is larger than that created by the latter. Thus, from the point of view of efficiency, reducing the tax on formal labor would have the highest return. That said, our model also indicates that the subsidy to informality deepens the economic distortion and, at the same time, puts additional pressure on the fiscal accounts; an important observation considering that resources for NCSI programs have increased noticeably over the last decade and, on present trends, will continue to do so.

From an economy-wide perspective, of course, what matters are not the effects of CSI or NCSI programs, but their joint impact (which, to the best of our knowledge, has not been measured previously). All in all, the impact of the CSI + NCSI tax-cum-subsidies on employment is substantial: salaried employment falls by 23% with its illegal component accounting for a 39% share; in parallel, employment in sector B is 45% higher. As seen, the effects on wages and the fiscal accounts are also large. Why? CSI taxes paid by firms and workers are 2.3% of GDP; this compares with 3.7% for the VAT. In addition, total subsidies to CSI and NCSI represent an additional 1.7% of GDP. Thus, taxes and subsidies worth 4% of GDP are being collected or channeled through the market for the most important non-traded input, labor; they are also being collected or channeled based on worker's labor contract (and, implicitly, firm’s size). It is not surprising that such a capricious mechanism to tax and subsidize, by distorting decisions of all firms and workers, has deep economic implications.

39 It is difficult to make a direct comparison of our results with those measuring the impact of individual NCSI programs. With that caveat, we note that, for instance, Bosch and Campos-Vazquez (2010) find that formal employment fell by 2.4% as a result of the Seguro Popular program, representing a loss of around 300,000 formal jobs. As shown in table A2.2, this program accounts for 24% of subsidies to all NCSI programs. Table 10, on the other hand, shows a 10% reduction in formal employment (or 1.36 million jobs) as a result of all NCSI programs. Clearly, one expects the impact of subsidies worth 1.25% of GDP to be larger than the impact of subsidies one-fourth that amount (although it is less clear that the relationship should be linear).
6.3 What are the fiscal costs of subsidies to CSI and NCSI?

Table 11 presents the results of an exercise where, starting from the status quo, government spending in SI is exogenously increased under two alternatives: first, a higher subsidy rate to CSI; second, increasing the per worker subsidy to NCSI. In both cases the increase is the same, arbitrarily set at 15,000 mp (or 0.12% of GDP). As expected, higher $\theta$ tilts the composition of employment towards formality while higher $\tau_{\text{NCSI}}$ does the opposite. In parallel, real wages and workers’ utility increase relatively more when CSI is subsidized, a result that follows from the fact that these subsidies reduce the distortion associated with the tax on formal employment, while subsidies to NCSI augment this distortion.

But the more interesting result centers on the impact on the fiscal balance, FB. Because higher subsidies to CSI increase formality, the tax base expands and revenues from VA and income taxes increase; as a result, even though spending goes up by 15,000 mp, the FB deteriorates only by 11,200 mp. On the other hand, higher subsidies to NCSI increase informality and contract the tax base with a negative impact on tax revenues; as a result, the FB deteriorates by more than 15,000 mp. The final outcome is this: the net cost of one peso of subsidies to CSI is 0.76 pesos vs. 1.03 pesos for subsidies for NCSI. This difference of 27% sharply illustrates the fact that the composition of spending in SI programs has substantive implications for the fiscal sustainability of SI policy.

Table 11: Net fiscal costs of CSI and NCSI subsidies

<table>
<thead>
<tr>
<th></th>
<th>Calibrated</th>
<th>Increase in CSI subsidies of 15,000 mp</th>
<th>Increase in NCSI subsidies of 15,000 mp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subsidies to CSI*</td>
<td>62.2</td>
<td>77.2</td>
<td>61.7</td>
</tr>
<tr>
<td>Subsidies to NCSI*</td>
<td>151.0</td>
<td>149.8</td>
<td>166.0</td>
</tr>
<tr>
<td>VAT Revenues*</td>
<td>457.9</td>
<td>459.9</td>
<td>456.9</td>
</tr>
<tr>
<td>IT Revenues*</td>
<td>392.4</td>
<td>392.7</td>
<td>392.3</td>
</tr>
<tr>
<td>Net impact on FB*</td>
<td>-</td>
<td>(-) 11.2</td>
<td>(-) 15.3</td>
</tr>
<tr>
<td>Net impact per peso</td>
<td>-</td>
<td>0.76</td>
<td>1.03</td>
</tr>
<tr>
<td>Formal employment**</td>
<td>12.29</td>
<td>12.50</td>
<td>12.17</td>
</tr>
<tr>
<td>Informal employment**</td>
<td>26.73</td>
<td>26.52</td>
<td>26.86</td>
</tr>
<tr>
<td>Formal wage***</td>
<td>1.000</td>
<td>1.005</td>
<td>1.002</td>
</tr>
<tr>
<td>Informal wage***</td>
<td>1.000</td>
<td>1.005</td>
<td>0.997</td>
</tr>
<tr>
<td>Worker’s utility***</td>
<td>1.000</td>
<td>1.005</td>
<td>1.003</td>
</tr>
</tbody>
</table>

*Thousands of million of pesos; **millions of workers; *** index.

6.4 What is the trade-off between VAT and CSI taxes?

Finally, we examine the impact on the real formal wage of different VA-CSI tax combinations. The left panel of Figure 1 plots an index of the real wage as a function of the
VAT rate on food and medicines given the $\tau^{CSI}$ rate and the evasion behavior of firms. We set this index at unity when both rates are set at their current value, i.e., $\tau^{VAT}_1 = 0$ and $\tau^{CSI}_1 = 0.38$; this is point A. The right panel plots the relationship between the real wage and the CSI rate given the same evasion behavior; it is downward sloping since CSI taxes are partly shifted to workers. The relationship is also drawn for $\tau^{VAT}_1 = 0$ and $\tau^{CSI}_1 = 0.38$; this is point A'.

We now carry out two sequential changes. First, we increase the VAT rate on food and medicines to 16%. On the left panel the real wage falls as workers face higher consumer prices; this is point B. Importantly, this impact is mediated by the extent of VAT compliance. Indeed, if the VAT on food and medicines was fully evaded by all firms, there would be no impact on consumer prices; conversely, under full compliance the impact on the real wage would be stronger. Our model indicates that with the observed compliance rates, when the VAT on food and medicines is set at 16% the real wage falls by 4%. On the right panel the increase in $\tau^{VAT}_1$ shifts the curve down, since for the same CSI rate the real wage is now lower; this is point B'. The second change consists in lowering $\tau^{CSI}_1$ to return the real wage to level it had before the change in $\tau^{VAT}_1$. On the right panel this is point C', while on the left it is point C.

The result is that the same real wage is obtained by the pair $(\tau^{VAT}_1, \tau^{CSI}_1) = (0,0.38)$ and by the pair $(\tau^{VAT}_1, \tau^{CSI}_1) = (0.16,0.30)$. The difference between the two is the point at which taxation is occurring. In the first, taxes are collected “at the door of the factory”; in the second case, taxation shifts partly to “the door of the store”. Of course, the revenues obtained in each case are different as the bases of these taxes are different (as are the possibilities of evading them). Further, under current Law the destiny of the revenues is different: taxes collected at the door of the factory can only be used for CSI benefits, while taxes collected at the door of the store can be freely disposed of (including paying for NCSI benefits and subsidizing CSI benefits). But these two differences are secondary. On one hand, the rates can be adjusted for given revenue targets; on the other, the destiny of the VAT can be earmarked by Law.

---

40 We set the VAT at 16% as the general VAT rate was raised from 15 to 16% in 2010. In addition, a rate of at least 16% is needed to finance the benefits of USI; see next section.
What is critical for our purposes is the reaction of firms and workers to these two changes. The VAT increase by itself will raise more revenues but will increase informality; the CSI rate reduction by itself will lower revenues but increase formality. The combination of the two is in principle ambiguous, and depends on how these rates are changed. But if revenues from the VAT could be earmarked for SI, this tax would achieve the same purpose as CSI taxes: changing the composition of worker’s consumption by lowering their disposable earnings and giving them in return SI benefits. The substantive differences would be three: (i) formal workers would pay less VAT (labeled CSI tax) at the door of the factory and more VAT (labeled VAT) at the door of the store; (ii) informal workers would pay more VAT at the door of the store; and, (iii) the base of the VAT would be wider than that of CSI taxes (as it includes profits and rents), and would not distort firms’ and workers’ decisions regarding salaried vs. non-salaried or legal vs. illegal contracts.

6.4 Assessment

We began this paper observing that redistribution and administrative convenience were the two main motivations behind Bismarck’s idea to fund SI out of wage-based taxes. Our results show, however, that in the case of Mexico CSI’s redistributive motivation is hardly achieved: CSI taxes are largely paid by workers. Moreover, this motivation results in: (i)
inequitable outcomes, as the wages of informal workers --who are not direct beneficiaries of CSI taxes-- are reduced; and (ii) unintended outcomes, as profits and quasi-rents in non-salaried activities increase. On the other hand, administrative convenience comes at a high price in terms of foregone revenues from income and VA taxes, and an equally high price in terms of economic efficiency, given the large distortions that CSI taxes introduce in the labor market. At the same time, even under full compliance, the coverage objectives of SI are only partly met, as non-salaried workers are excluded. When attempts are made to remedy this situation through NCSI programs, the redistribution and administrative motivations are lost, as these programs are funded completely from other sources of revenues, while efficiency is further punished and fiscal costs augmented given higher evasion.

The net result is a very distorting system for financing SI and very erratic coverage against risks, in a context of large mobility of workers between formal and informal status. Given subsidies to CSI and the actual incidence of CSI taxes, when a Mexican worker is formally employed, he receives a bundled package of SI benefits that is 16% subsidized by the government (from VAT, other taxes or oil rents), 54% (= 0.84*0.64) by himself and 30% (=0.84*0.36) by the firm that hires him (since as seen 64% of CSI taxes are shifted to workers). While he holds this status, he is obligated to contribute to the revenues of state governments given their tax on salaried labor; to save for his retirement and for a house; to purchase health, disability and work-risk insurance and day care services; and is protected by severance payments in case of job loss. However, when the same worker is informally employed, he can benefit from an unbundled package of SI programs that are 100% subsidized by the government. While he holds this status, he does not have to pay any state labor tax; he can voluntarily access a health insurance program; eventually receive a pension without having saved for one; maybe access a day care center or receive a housing subsidy; but be left without any severance pay in case of job loss. Firms, in turn, face different obligations depending on whether they hire workers under one labor contract or another, inducing them into illegal behavior and socially inefficient choices. And the fisc, finally, loses revenue given a narrowed tax base.

This state of affairs would be amply justified, of course, if there were no other alternatives for the government to provide SI to workers, as the option of leaving them unprotected against risks is not an option. The relevant question in this context is thus not whether SI should be funded from VA taxes; as shown, taxes at the door of the store (and
other sources of revenue) already fund a 100% of NCSI and 16% of CSI. Moreover, it is inevitable that this be so because there will always be non-salaried contractual relations in the economy and, given imperfect enforcement of the tax on salaried labor, illegal salaried workers. Rather, in a context where the government wants formal and informal workers to be covered by SI, the relevant question, in our view, is twofold. Does it make sense to continue to provide different SI benefits to the same worker depending on his labor contract, and to fund these benefits with a mix of labor and other taxes? Is there an alternative mix of financing that produces fewer productivity-reducing distortions, enlarges the tax base, and –critically— is able to offer the same protection against risks that are common to all workers?

7. Universal social insurance and the end of informality

7.1 What does USI consist of?

This and the next section describe our proposal for USI. This section assumes that there are a set of risks against which all workers need to be protected regardless of labor status, namely, longevity, illness, disability and death. We separate these risks from others that workers may face but that are specific to their salaried status such as being fired by their boss or suffering an accident at work because the firm fails to comply with safety standards. These latter risks are dealt with in the next section, where we also propose complementary pension benefits for salaried workers. We highlight that when account is taken of these complementary pension benefits, our proposal implies that the health, life and disability and work-risk insurance as well as retirement pensions that formal workers currently receive under CSI would remain the same under USI.

For the purposes of this section we define universal social insurance, USI, as the provision of health, life and disability insurance and retirement pensions of the same level of quality (health) or monetary payment (pensions) to all workers or their families. We denote the per worker monetary cost of those benefits as $\tau^{USI}$. More precisely, we assume that under USI all workers are entitled to: (i) the same health benefits currently provided by CSI when they are formally employed; this is the IMSS health program, whose coverage is broader than that provided by NCSI health programs when they are informally employed, (ii) a contribution for life, disability and retirement pensions equal to that received by a worker earning two minimum wages, and (iii) a guarantee of a minimum pension of one minimum wage if workers,
as is currently the case under CSI, are registered with USI for at least 25 years. (In the next section we extend USI to include additional pension benefits.) We exclude housing and day care services from USI.42

The benefits of USI are delivered in a similar fashion as those presently provided by CSI. Retirement pensions would operate as a defined contribution program with monthly contributions deposited in workers’ individual retirement accounts in the Afore of their choice, and with annuities bought from private insurance companies at the time of retirement; in turn, for risk pooling purposes, contributions for death and disability insurance would be deposited in the same common reserve fund managed by IMSS as currently occurs under CSI, with payouts for permanent disability or survivorship pensions also taking the form of annuities. Thus, very importantly, the pension components of USI are fully funded, with no contingent liabilities. In parallel, health benefits would continue to be provided by IMSS and state governments. In particular, IMSS would receive the same payment for health per enrolled worker as under CSI, while state governments would receive a higher per worker payment vis-à-vis what they currently get from the NCSI health programs (equal to that received by IMSS).

Appendix 3 provides details of how $\tau^{USI}$ was calculated; suffice it to say here that it equals 14,330 pesos per worker annually, which compares with subsidies for CSI and NCSI of 5,062 and 5,652 pesos, respectively. Of this total, 70.6% corresponds to health benefits, 24.3% to retirement pensions, and 5% to life and disability insurance. Very importantly, this benefit level implies more than doubling public subsidies for SI, from 1.7% of GDP under CSI + NCSI, to 4.4% under USI; clearly, a major change.43

42 Mexico is one of the few countries in the world to include housing and day care as obligatory components of CSI. These needs are of a different nature than the risks traditionally associated with SI; see Levy (2008). Our calculations exclude the resources necessary to provide housing and day care benefits to all workers. Of course, the government could continue to subsidize housing and day care programs, but these would no longer be considered part of SI. Our calculations also assume that state taxes on salaried labor are also eliminated, and that state governments are compensated by the federal government for the foregone revenues; see below.

43 The proposal also implies increasing total national spending in SI. Under CSI + NCSI this spending is 3.9% of GDP (excluding SI spending for public sector workers and the transition costs of the change from the pay-as-you-go to the defined-contribution scheme resulting from the 1997 pension reform). Of this total, 1.7% are government subsidies (for CSI and NCSI), and 2.2% CSI taxes paid by firms and formal workers. Under USI public spending in SI is, as noted, 4.4% of GDP. As discussed in the next section, however, to this sum one needs to add complementary pension benefits paid by firms and salaried workers worth 0.4% of GDP. Thus, all in all, spending in SI would increase from 3.9 to 4.8% of GDP. There would also be a change in composition as public spending in day care and housing would disappear (by the equivalent of 0.5% of GDP), with spending in health and pensions therefore increasing by 1.4% of GDP.
7.2 The labor market and the fiscal balance under USI

There are three critical implications of USI. First, CSI and NCSI benefits are eliminated and replaced by USI benefits, so worker’s utility is:

\[ U_f = w_f + \beta^{USI} \tau^{USI} \quad ; \quad (24) \]

\[ U_i = w_i + \beta^{USI} \tau^{USI} \quad ; \quad (25) \]

where \( \beta^{USI} \in [0,1] \) refers to workers valuation of the bundle of USI benefits. As a result, there are no incentives for workers to seek one form of employment vs. another just on the basis of differences in SI benefits.

Second, the cost of labor is the same regardless of the nature of the labor contract. In terms of table 5 this implies, critically, removing all taxes on formal employment and subsidies to informal employment. As a result, firms have no incentive to offer workers one type of contract vs. another, to change their size as a strategy to avoid CSI taxes, to mask a salaried contract as non-salaried, or to out-source activities or rotate workers just to evade SI regulations. Differently put, USI implies eliminating all distortions in firms’ and workers’ behavior in the labor market stemming from SI policy, a situation that would have an unambiguously positive impact on productivity.\(^{44}\)

The third implication, finally, is that an ear-marked consumption tax is needed to cover the costs of USI, a tax that has the same base, method of accreditation and enforcement technology as the VAT. Indeed, were it not for the fact that the proceeds of this tax are ear-marked for USI, it would be exactly the VAT. Because operationally this tax is collected along with the non ear-marked VAT, we can divide the VAT rate in two parts:

\[ \tau^{VAT} = \alpha \tau^{VAT} + (1-\alpha) \tau^{VAT/USI} = \tau^{VAT/USI} + \tau^{VAT/GP} \quad ; \quad (26) \]

where \( \alpha \) is the share of the total VAT rate ear-marked for USI, \( \tau^{VAT/USI} \), with the remainder being the share used for general purposes, \( \tau^{VAT/GP} \). Under our proposal, \( \tau^{VAT} \) is obtained by equalizing the VAT rate on \( I_1 \) (food and medicines) and \( I_2 \) (all other goods) and setting both at

\(^{44}\) An important exception is the distortions created by severance pay which so far have been excluded from the analysis; see the discussion at the end of section 8.
16%.\textsuperscript{45} In turn, total VAT revenues, $R^{\text{VAT}}(\ldots)$, are separated into revenues ear-marked for USI, $R^{\text{VAT/USI}}(\ldots) = \alpha R^{\text{VAT}}(\ldots)$, and non-earmarked revenues $R^{\text{VAT/GP}}(\ldots) = (1 - \alpha)R^{\text{VAT}}(\ldots)$.

Very importantly, from a legal perspective $\tau^{\text{VAT/USI}}$ is labeled as “Contribution for USI”, implying that proceeds from this contribution are wholly ear-marked for SI. In turn, $R^{\text{VAT/USI}}(\ldots)$ is registered in a separate account of the federal government’s budget, is not subject to revenue-sharing with the states, and is divided in three separate sub-accounts with no possibilities of transfers between them, so as to clearly identify and manage resources for health, life and disability pensions, and retirement pensions. Resources from the health sub-account are transferred to IMSS and state governments to provide health services of equal quality to all workers; resources from the life and disability insurance sub-account are transferred to the common IMSS-managed reserve fund; and resources from the retirement pension sub-account are transferred on a per worker basis to their individual retirement accounts.

A critical condition for the fiscal sustainability of USI is that:

\begin{equation}
(27) \quad R^{\text{VAT/USI}}(\ldots, \tau^{\text{VAT/USI}}, \ldots) = \tau^{\text{USI}} L
\end{equation}

Equation (27) is central: it states that the costs of USI, given by the exogenously chosen level of benefits and the size of the labor force, must be fully financed from the share of VAT revenues legally ear-marked for USI. This implies, as noted, that $R^{\text{VAT/USI}}$ cannot be used for purposes other than SI but, conversely, that USI must be fully funded from these revenues; the government could not transfer resources from other sources (including oil rents!) to fund SI. Critically, therefore, $\tau^{\text{USI}}$ could not be set independently of $\tau^{\text{VAT/USI}}$, and any increases to $\tau^{\text{USI}}$ that are deemed socially desirable would have to be funded with additional contributions.

The government’s fiscal balance under USI is therefore written as:

\begin{equation}
(28) \quad FB = [\bar{R} + R^{\text{VAT/USI}}(\ldots) + R^{\text{VAT/GP}}(\ldots) + R^{\text{IT}}(\ldots)] - [\bar{G} + \tau^{\text{USI}} L]
\end{equation}

Comparing this with the fiscal balance under the CSI + NCSI configuration given by equation (22), and considering the constraint imposed by (27), it is obvious that this particular method of funding SI has two major advantages: (i) it ensures the long term fiscal sustainability of

\textsuperscript{45} Our proposal implies as well raising VAT rates in the border to 16% (from 11%), although this is not captured by our model since there is no regional dimension.
Mexico’s SI policy, and (ii) it makes the effort made by society to fund SI fully transparent, tightly linking benefits with contributions.46

In sum, we propose a social-cum-fiscal reform that provides all workers with the same SI benefits; that eliminates subsidies to non-salaried labor and taxes to salaried labor while changing the burden of taxation towards consumption; that explicitly links contributions with benefits; and that creates no contingent liabilities and puts no pressure on the government’s fiscal balance. This proposal, of course, needs to be evaluated in a context where firms’ can evade the VAT; where there is a large number of family firms and self-employed workers (sector B) which, although in principle contributing under the Repeco regime, are by-and-large beyond the reach of the tax authority; and where compensations need to be offered to poor households to fully offset any negative effects of the VAT increase.

7.3 Fiscal reform under CSI and NCSI vs. USI

Table 12 shows the impact a VAT reform that homogenizes all rates at 16% under three scenarios: one with the current configuration of CSI and NCSI taxes and subsidies; one where this configuration is replaced by USI but no compensations are offered to anyone; and one where compensations are included as part of the fiscal costs of USI. We focus on the impact of the reform on the fiscal balance (panel A), wages and utility (panel B), and employment (panel C).

Consider first the VAT reform in the absence of any changes to SI policy. As expected, there is an important increase in VAT revenues, of approximately 3% of GDP (although there is a loss of revenue from income taxes as firms pay more VAT). As expected as well, the real wage falls given higher consumer prices. But, very importantly, note that the reform increases the level of informality in the economy, as employment in the non-taxed sector B expands and as firms producing $I_1$ and $I_2$ increase the share of salaried workers hired illegally. This increases spending in subsidies to NCSI (and reduces it in subsidies for CSI), with a net increase in spending (given (23)). Because the VAT reform is centered in food and medicines ($I_1$), this

46 Note that something akin to equation (27) already holds for CSI, given Mexico’s defined-contribution retirement system. Indeed, it is the existence of NCSI programs that breaks the link between contributions and SI benefits, and has opened the door for unfunded commitments and new contingent liabilities derived from health or pension programs. Our proposal re-establishes that link and closes the door to future unfunded commitments. The political economy implications of this are large, as neither the President nor Congress could offer additional SI benefits to workers without in parallel assuring the source of funding. This is already the case for CSI programs, but not for NCSI programs. See Lindert (2004) for a historical discussion of the fiscal sustainability of SI policy.
sector contracts while there is a marginal increase in the already taxed sector \( (I_2) \); indeed, following expression (13), as the price of \( I_1 \) in terms of \( I_2 \) increases the composition of good A tilts towards \( I_2 \). But the biggest gainer is the non-taxed sector \( B \); evidently, as the average VAT rate in sector \( A \) increases the non-taxed sector of the economy absorbs more resources. On the other hand, when the effects of these spending and revenue changes are factored into the fiscal balance, the result is a surplus of 1.32\% of GDP, representing a turnaround of 2.8\% of GDP from the pre-VAT reform situation (since in this scenario none of the additional revenues are spent).

Table 12: Fiscal reform for USI: \( r_{1,\text{VAT}} = r_{2,\text{VAT}} = 0.16 \)

<table>
<thead>
<tr>
<th>Panel A*</th>
<th>Calibrated</th>
<th>VAT reform under CSI + NCSI</th>
<th>VAT Reform under USI</th>
<th>VAT reform, USI &amp; compensations</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAT Revenues</td>
<td>457.9</td>
<td>824.2</td>
<td>873.6</td>
<td>873.6</td>
</tr>
<tr>
<td>(% GDP)</td>
<td>3.76</td>
<td>6.79</td>
<td>7.1</td>
<td>7.1</td>
</tr>
<tr>
<td>IT Revenues</td>
<td>392.4</td>
<td>377.3</td>
<td>383.0</td>
<td>383.0</td>
</tr>
<tr>
<td>Subsidies to CSI</td>
<td>62.2</td>
<td>58.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Subsidies to NCSI</td>
<td>151.0</td>
<td>154.6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Subsidies to USI</td>
<td>0</td>
<td>0</td>
<td>559.3</td>
<td>559.3</td>
</tr>
<tr>
<td>Compensations to the poor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28.8</td>
</tr>
<tr>
<td>Compensations to states</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26.0</td>
</tr>
<tr>
<td>IMSS pension liabilities</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27.0</td>
</tr>
<tr>
<td>Fiscal balance</td>
<td>(-) 190.4</td>
<td>160.7</td>
<td>(-) 148.5</td>
<td>(-) 230.3</td>
</tr>
<tr>
<td>(% GDP)</td>
<td>(-) 1.56</td>
<td>1.32</td>
<td>(-) 1.21</td>
<td>(-) 1.88</td>
</tr>
</tbody>
</table>

| Panel B*** | | | |
| CPI | 1.000 | 1.026 | 1.027 | 1.027 |
| Real wage | 1.000 | 0.965 | 1.158 | 1.158 |
| Worker’s utility | 1.000 | 0.991 | 1.203 | 1.203 |

| Panel C*** | | |
| Employment I_1 | 5.24 | 4.63 | 6.06 | 6.06 |
| Employment sector B | 18.93 | 19.42 | 13.41 | 13.41 |
| Legal salaried employment | 12.29 | 11.66 | 25.61 | 25.61 |
| Illegal salaried employment | 7.80 | 7.94 | 0 | 0 |

*panel A, thousands of millions of pesos; **panel B, index; ***panel C, millions of workers.

The fiscal-cum-SI policy reform produces a very different outcome. First, for the same increase in the VAT rate there is an extra gain in revenues of approximately 0.3\% of GDP; there is also a small gain in income taxes. Since enforcement efforts are assumed constant, these gains derive only from the fact that, despite the VAT change, the level of informality in the economy is reduced as the tax on formality and the subsidy to informality stemming from

---

47 It is easy to see that if the additional revenues produced by the VAT reform were spent in NCSI programs, the increase in informality would be more pronounced. This is a relevant point as in Mexico strengthening NCSI programs has been used as a justification to increase the VAT (as occurred in the 2010 Budget Proposal).
SI policy are removed. Indeed, salaried employment increases by 27%, illegal salaried employment disappears (as there is no tax to evade), and non-salaried employment in sector B contracts since the supply of labor to this sector falls.  

Second, note that despite the VAT increase, employment in sectors I₁ and I₂ increases. The gain in I₂ is expected: firms pay the same VAT rate as before but now face lower labor taxes, so clearly output and employment expand while the incentives to evade fall. The gain in I₁ is more interesting: on one hand, this sector now pays VAT so *ceteris paribus* one expects it to contract and evade more (as indeed is the case, as shown in column 2 of table 12); on the other hand, when the tax on formal labor is removed one expects the sector to expand and to evade less. What is noteworthy is that the net effect is positive. Contrasting this result with the one obtained when only the VAT is increased highlights the importance of the *simultaneity* of the fiscal and social reform, a point that from the point of view of policy is of the essence. The net loser, unsurprisingly, is sector B; a result that in turn highlights that the root cause of the large sector of self-employment and family firms observed at present in Mexico is the CSI + NCSI configuration, not the VAT. The net impact of these effects is to expand the tax base, generating the extra 0.3% of GDP in VAT revenues vis-à-vis the reform under the CSI + NCSI configuration.

Third, the reform produces an increase in the real wage despite higher VAT taxes. This result is, in part, the other side of the coin of the fact that CSI taxes are in large part shifted to workers: their removal is reflected in a higher wage. In part, however, it also reflects that eliminating the CSI tax results in a more efficient labor allocation as firms’ labor costs fall. In addition, eliminating NCSI subsidies makes employment in sector B less profitable and increases the supply of labor to the segment of the economy where its marginal revenue product is highest, i.e., to sectors I₁, I₂ and A. Thus, the reduction in the size of the non-taxable sector B has two effects: as noted, it expands the tax base; but it also contributes to increase the average productivity of labor.

But most important for our purposes, fourth, note that ignoring any compensations, the social-cum-fiscal reform *reduces* the fiscal deficit by 0.35% of GDP vis-à-vis the status quo, given the level at which USI benefits are set. These benefits represent 64% of total VAT

---

48 If, more realistically, sector B also required intermediate inputs from sector I₁, sector B would contract more as, for the reasons discussed earlier, input prices would increase without family firms or self-employed workers having the possibility of receiving a credit for the higher VAT.
revenues, suggesting in turn that $r_{\text{VAT/USI}}$ and $r_{\text{VAT/GP}}$ are in the order of 10 and 6%, respectively; see equation (26).

The last column of table 12, finally, includes as part of the fiscal costs of the proposal three additional costs: (i) direct transfers to the poor to offset the negative income effect of the VAT increases, as discussed in section 9; (ii) compensating state governments for the foregone revenue from eliminating state taxes on salaried workers, and (iii) servicing the pension liabilities of IMSS-own workers directly by the federal government (as explained in Appendix 3). The relevant point to notice is that when these costs are included, the net fiscal impact changes from a reduction in the fiscal deficit of 0.35% of GDP vs. the status-quo, as noted above, to an additional deficit of 0.32% of GDP.

We could present results with a higher VAT rate or with lower USI benefits so as to leave the fiscal balance unchanged. But that is not the main point here. Rather, the point is that our calculations show that a proposal for USI financed from a consumption tax need not represent an unmanageable burden on the fiscal balance. Of course, whether this is so or not depends on considerations that escape the scope of our paper, having to do with societal choices for taxation and other claims on public resources. That said, it is useful to be clear about the real policy alternatives, as the relevant comparison is in all likelihood not between USI and the 2008 status quo. The more relevant comparison is between USI and the observed trends in SI policy. In particular, if the Mexican government continues to respond as it has done so far to societal demands for increased social equity and improved coverage of risks for

---

49 Given the already high vertical imbalance and dependence of state governments on transfers from the federal government, it would be better if the revenues foregone by state government were replaced from another source like a state gasoline tax or improved incentives to collect in the Repeco regime, as opposed to being compensated by the federal government. This would reduce the costs of compensations by 0.2% of GDP.

50 As noted, our calculations ignore the costs of maintaining subsidies for housing and day care programs, for approximately 0.5% of GDP (of which housing represents 0.4%). Assuming these are maintained along with all compensations, the fiscal deficit would increase relative to the status quo by 0.82% of GDP (or 0.62% excluding compensations to state governments).

51 A scenario where the general VAT rate is set at 17% produces about 50,000 mdp of additional revenues, or 0.4% of GDP. This would reduce the fiscal deficit vs. the status quo to 0.1% of GDP after considering all compensations (or to 0.3% if compensations to state governments are excluded). Alternatively, if housing and day care programs are maintained, the fiscal deficit vs. the status quo would increase by about 0.4% of GDP (or 0.2% if compensations to states are excluded). This suggests setting the VAT rate somewhere between 16 and 17%. On the other hand, our proposal would also have important implications for personal income taxes that we do not model here, and which could likely generate more revenues as salaried employment expands, or certain exemptions eliminated (e.g., deductions for health insurance) given a context where the government provides all workers with SI benefits.
informal workers through further growth in NCSI programs in parallel to existing CSI programs, the result will be a narrower tax base, persistent productivity losses, increased incentives to informality, and further de-linking of benefits from contributions.

7.4 The end of informality?

Fiscal and efficiency considerations aside, the most important effect of our proposal is to provide all workers with the same coverage against risks. If, following Kanbur (2009) and the discussion of section 2, informality is defined with respect to the un-observance of a particular regulation –SI coverage–, the reform indeed ends informality: all workers are covered by the same SI.

Completing Mexico’s truncated Welfare State, i.e., replacing CSI + NCSI with USI, helps solve critical problems present in a labor market characterized by high frequency of transits across forms of employment. One, workers receive the IMSS health package regardless of their labor contract, with positive implications for their own welfare (and large implications for public subsidies for health which, as shown in Appendix 3, would increase by 52%, or an additional 1.1% of GDP). Two, the adverse selection problem created by the co-existence of two health insurance systems, one free and voluntary and the other costly and obligatory, are eliminated; health risks are pooled more effectively. Three, workers contribute to their retirement pension during their entire working life, not only when they are formally employed; as a result, as elaborated on in the next section, retirement pensions would be more generous. Four, all workers could qualify for the minimum pension guarantee because they would accumulate for their retirement pension throughout their working life. This feature is particularly relevant for low wage workers, the majority of whom, as a result of high transits between formal and informal status, will not benefit from this guarantee (despite being the intended beneficiaries). More generally, five, workers would be protected against risks regardless of the vagaries of their labor contract, increasing the efficacy of insurance. Finally, six, one would expect that because CSI taxes are eliminated, contractual changes and worker rotations that result from firms’ strategies to evade these taxes would cease, facilitating more

52 As noted, during 1997-2006 the average contribution density for low wage workers was 49%, implying the need to work 50 years to benefit from this guarantee (since under CSI 25 years of formality are needed to qualify). In contrast, under USI non-salaried employment would also be considered (implying contribution densities of 100%).
durable relations between firms and workers, more on-the-job learning and increased investments in labor training.

But the end of informality is not the end of illegality. In our model, after the reform firms producing $I_1$, $I_2$ and $A$ continue to evade VA and income taxes, and a sector beyond the reach of the tax authorities continues to exist, sector B, although diminished in size. Our reform proposal does not solve the difficulties of imperfect enforcement, nor does it directly affect other determinants of the size distribution of firms that in turn have an impact on enforcement (like access to credit). Further, our proposal fails to address issues associated with the high costs of registration and compliance with sanitary, environmental and related regulations, which would also need to be tackled as part of a larger effort to reduce illegal behavior.

That said, there are additional potentially positive effects of our proposal not captured by our model. On one hand, unifying the VAT rates would significantly simplify the administration of this tax, as the current cumbersome system of exemptions by location and type of good—and associated possibilities for arbitrage and graft—would be unneeded; this should in principle facilitate enforcement. On the other, recall that our model takes $f(K_1)$ and $f(K_2)$--the distribution of the capital stock in each sector—as given. In this context, we speculate that as the implicit tax on firm size associated with the CSI + NCSI configuration is removed, the profitability of larger vis-à-vis smaller firms would increase changing the size distribution of firms towards larger firms, and this itself would be a factor facilitating enforcement. Exploring these issues, however, is left for further research.

**7.5 Decomposition of the Increase in VAT Revenues**

To shed further light into our results, table 13 divides total VAT collections into those originating from goods $I_1$, $I_2$ and $A$, and provides the estimated values of the aggregate rates of compliance for each of the VAT reform scenarios (see equation (10) and Appendix 1).
Table 13: Decomposition of the Total Increase in VAT Revenues

<table>
<thead>
<tr>
<th></th>
<th>Calibrated</th>
<th>( \text{VAT} = 15% ), ( \frac{n_{VAT}}{\xi_{VAT}} = 0 ), CSI + NCSI</th>
<th>( \text{VAT} = 16% ), ( \frac{n_{VAT}}{\xi_{VAT}} = 0 ), CSI + NCSI</th>
<th>( \text{VAT} = 16% ), ( \frac{n_{VAT}}{\xi_{VAT}} = 0.36 ), CSI + NCSI</th>
<th>( \text{VAT} = 16% ), ( \frac{n_{VAT}}{\xi_{VAT}} = 0.36 ), USI</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAT Revenues*</td>
<td>457.9</td>
<td>683.7</td>
<td>726.6</td>
<td>824.2</td>
<td>873.6</td>
</tr>
<tr>
<td>( \xi_{VAT} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36.0</td>
<td>42.8</td>
</tr>
<tr>
<td>( \xi_{VAT} )</td>
<td>211.6</td>
<td>211.6</td>
<td>222.2</td>
<td>224.8</td>
<td>267.4</td>
</tr>
<tr>
<td>( \xi_{VAT} )</td>
<td>246.3</td>
<td>472.1</td>
<td>504.4</td>
<td>563.4</td>
<td>563.4</td>
</tr>
</tbody>
</table>

 Compliance Rates

<table>
<thead>
<tr>
<th>( \frac{n_{VAT}}{\xi_{VAT}} )</th>
<th>0.00</th>
<th>0.00</th>
<th>0.00</th>
<th>0.36</th>
<th>0.36</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{n_{VAT}}{\xi_{VAT}} )</td>
<td>0.61</td>
<td>0.61</td>
<td>0.61</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>( \frac{n_{VAT}}{\xi_{VAT}} )</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.53</td>
<td>0.53</td>
</tr>
</tbody>
</table>

*Thousands of million of pesos.

The calibrated equilibrium in the second column shows, as expected, that the compliance rate in \( \xi_{1} \) is zero when \( \tau_{1}^{VAT} = 0 \). VAT collections all come from \( \xi_{2} \), where the VAT rate is 15\% and the rate of compliance 0.61; and from final good A, where the VAT rate and the rate of compliance is a weighted average of the corresponding rates in sectors \( \xi_{1} \) and \( \xi_{2} \). Column three measures the impact of the VAT reform when \( \tau_{1}^{VAT} = 15\% \), while column four sets the VAT rate in both sectors \( \xi_{1} \) and \( \xi_{2} \) at 16\%. In these two columns we assume that the VAT reforms occurs in isolation (i.e., the CSI + NCSI configuration is maintained).

Now, to identify the impact that compliance in intermediate sector \( \xi_{1} \) has on compliance in the final sector A, in columns three and four we artificially assume that, despite the fact that all firms in sector \( \xi_{1} \), now face a positive VAT rate, none comply with the VAT (clearly not a profit maximizing point). As a result, in both these columns the compliance rate in sector A is the same as before. Under this assumption, a VAT reform where rates are set at 15\% produces total VAT revenues of 683,700 mp (column three); and of 726,600 mp when these rates are set at 16\% (column four). Note that in both cases most of the additional revenues come from final good A since, by assumption, sector \( \xi_{1} \) pays no additional VAT (and sector \( \xi_{2} \) pays a VAT rate that is only one percentage point higher in column four).

The interesting result arises in column five where, while still maintaining the CSI + NCSI configuration, we now allow sector’s \( \xi_{1} \) aggregate compliance rate to be determined by equation (10) which, given the size distribution of firms in that sector, is 0.36. Two effects follow. First, firms in that sector now pay 36,000 mp of VAT, compared to none before. Second, VAT paid by sector A increases by 59,000 mp, with the net result that total VAT
revenues increase by 13% (from 726,600 to 824,200 mp or by 0.8% of GDP), for the same 16% VAT rates. These two effects, particularly the revenue increase in sector A, sharply illustrate, in accordance with our earlier discussion and with De Paula and Scheinkman’s (2010) “chain effects”, that higher compliance in the intermediate sector induces higher compliance in the final sector. Indeed, sector’s A compliance rate increases from 0.43 to 0.53. Differently put, reduced evasion in I₁ translates into reduced evasion in A. The chain of accreditation and payment is strengthened: because firms in I₁ pay more VAT, they issue more VAT invoices, which in turn induces firms in sector A to pay more VAT to be able to get credit for the VAT paid to firms in I₁.

We highlight that these “chain effects” are not small: the increase in revenues in sector A from having sector I₁ pay some VAT translates into 59,000 mp of additional revenues, or approximately 0.4% of GDP. This compares with the direct VAT payments of sector I₁ of 36,000 mp (or 0.3% of GDP). Thus, “chain effects” matter greatly for policy: they show that focusing only on the additional VAT that firms in sector I₁ would pay by eliminating the exemptions to food and medicines misses the significant fact that, by completing the value added chain, the higher rate on I₁ induces greater compliance with the VAT throughout the whole economy.

Finally, the last column in table 13 picks up the effect of the SI reform, from CSI + NCSI to USI. As discussed, firms in sectors I₁ and I₂ face lower labor costs, expand employment and output and, for the same compliance rates, pay more VAT. The net result is that the increase in total VAT revenues from 457,900 mp (or 3.7% of GDP) under the status quo to 873,600 mp (or 7.1% of GDP) under our proposal, results from: (i) an increase in VAT in I₁ from 0 to 15%, as captured in column three vs. column two; (ii) an increase in VAT from 15 to 16% for I₁ and I₂, as captured in column four vs. three; (iii) compliance in I₁ and its impact on compliance in sector A, as captured in column five vs. four; and (iv) reforming the SI regime, as captured in column six vs. five.

---

53 In parallel, note that the aggregate compliance rate in I₁ is lower than in I₂, 0.61 vs. 0.31, a result that follows from the fact that the size distribution of firms in sector I₁, where food and medicines are produced, is skewed to the left in comparison to that of sector I₂. This can be verified by noting that a larger share of employment in sector I₁ occurs in small firms vs. sector I₂ (see table 8).
Are VAT collections of 7.1% of GDP feasible?

We end this section with a brief discussion of the robustness of our results. Because the costs of USI benefits are exogenously given, the additional revenue generated by the VAT reform is the central question to determine the fiscal feasibility of our proposal.

Table 14 compares the VAT to GDP ratio observed in ten Latin American countries with three ratios for Mexico: the one observed in 2008 with a 15% general rate but exemptions to food and medicines; the one predicted by our model when rates are homogenized at 16% but CSI + NCSI taxes and subsidies remain; and the one predicted by our model with the same reform but with USI replacing the CSI + NCSI configuration. Rates and revenue ratios vary given country differences in exemption regimes, shares of exports and imports in GDP, evasion and informality (as reflected in the “productivity” measure obtained from dividing the revenue ratio over the basic rate).

Table 14: Rates, revenues and productivity of VAT

<table>
<thead>
<tr>
<th>Basic Rate (%)</th>
<th>Revenues/GDP</th>
<th>Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSI + NCSI and exemptions (2008)</td>
<td>15*</td>
<td>3.8</td>
</tr>
<tr>
<td>CSI + NCSI no exceptions</td>
<td>16</td>
<td>6.8</td>
</tr>
<tr>
<td>USI no exceptions</td>
<td>16</td>
<td>7.1</td>
</tr>
<tr>
<td>Uruguay</td>
<td>22</td>
<td>10.7</td>
</tr>
<tr>
<td>Brasil</td>
<td>17</td>
<td>8.3</td>
</tr>
<tr>
<td>Chile</td>
<td>19</td>
<td>7.4</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>15</td>
<td>7.3</td>
</tr>
<tr>
<td>Argentina</td>
<td>21</td>
<td>7.2</td>
</tr>
<tr>
<td>El Salvador</td>
<td>13</td>
<td>7.0</td>
</tr>
<tr>
<td>Bolivia</td>
<td>13</td>
<td>6.4</td>
</tr>
<tr>
<td>Venezuela</td>
<td>12</td>
<td>6.4</td>
</tr>
<tr>
<td>Colombia</td>
<td>16</td>
<td>5.5</td>
</tr>
<tr>
<td>Guatemala</td>
<td>12</td>
<td>5.4</td>
</tr>
</tbody>
</table>

*The average rate given exemptions is close to 10%. Source: Own calculations from countries data.

We make four points: first, by Latin American standards Mexico is currently an outlier in terms of its very low revenue ratio. Second, a uniform 16% rate is not higher than that observed in Uruguay, Brazil, Chile, Argentina and Colombia. Third, our model prediction of a revenue ratio of 7.1 is within the range of experiences observed in Latin America, particularly if consideration is made that this ratio would derive from a reduction in the size of the non-taxed sector and of firms’ incentives to evade.

*In 2007 VAT rates for OECD countries, of which Mexico is a member, were: Belgium, 21; Finland, 22; France, 19; Greece, 18; Italy, 20; United Kingdom, 17.5; Portugal, 17; Spain, 16; Sweden, 25; see Bird and Gendron (2007). More recently some of these countries have raised their rates. On the other hand, Hernández (2009) compares Mexico’s performance with countries like Canada and Korea who also have important exemptions and special treatments, and finds that VAT revenues in Mexico are low.
The fourth point follows from comparing the two simulated VAT to GDP ratios. In our model, changes in VAT revenues result from labor reallocations as firms and workers respond to changes in VA and SI tax and subsidy rates. As discussed, the VAT reform that maintains CSI + NCSI taxes and subsidies re-allocates labor in the direction of informality, eroding the tax base. This suggests that if USI replaced the CSI + NCSI configuration in parallel with the VAT reform, VAT revenues would be at least 6.8% of GDP. Indeed, this figure would result from eliminating the tax on formal labor and the subsidy to informal labor, but with firms demanding less legal salaried workers, while illegal salaried employment, self-employment and employment in family firms expands in response! Differently put, even if the elasticities of the demand for labor in all firms and sectors were not only different in value from the ones implied by our model, but of the opposite sign, the lower bound for the VAT/GDP ratio would be 6.8%. Of course, we expect labor to flow in the opposite direction when USI is introduced generating, according to our model, an additional 0.3% of GDP in VAT revenues. From this perspective and in the light of the comparison of Mexico vis-à-vis other countries in Latin America, it seems reasonable to expect VAT revenues in the order of 7.1% of GDP in response to the proposed fiscal-cum-SI policy reform.55

8. Complementary social insurance for salaried workers

This section extends our proposal to cover risks that are specific to salaried status and to allow for pensions that are proportional to workers’ wages.

We begin with retirement pensions. In a context where there are wage differences across workers (due to, say, skill differences), our proposal so far implies that workers who earn less than two minimum wages would save a higher share of their earnings than what workers in that earnings range save under CSI (since under USI all workers receive a contribution to their individual retirement account equal to that received by a worker earning two minimum wages); workers earning two minimum wages would save the same share as under CSI; and workers earning more than two minimum wages would save a lower share. Therefore, retirement pensions would also be the same, implying declining replacement rates as earnings increase. However, some arguments suggest that retirement pensions should also help workers smooth consumption between their active time and their time as retirees. To

55 Indeed, over the medium term one would expect even higher VAT revenues as the marginal product of capital in the taxable sector of the economy increases with the reform (because salaried employment expands), while it falls in the non-taxable sector (because the subsidy to non-salaried employment disappears).
achieve constant replacement rates, contributions need to be set as a share of individual worker’s earnings and not as the same amount for all workers.

We add consumption smoothing to our proposal through a two-pillar retirement pensions system: one, a fixed contribution to workers’ individual retirement accounts regardless of their wages as described in the previous section; two, a variable contribution to the same accounts that depends on worker’s wages. We elaborate on this proposal immediately below but point out a critical limitation up-front. Savings for the second pillar would only occur when workers are in salaried employment, because only then could their wages be observed and taxed “at the door of the factory”, in the same way CSI taxes currently operate. For the reasons discussed earlier in this paper, self-employed workers and workers in family firms could not be taxed for the second pillar, and the government would be unable to smooth their consumption through time.\(^5\)

We set the contribution rate for the second pillar as a function of salaried workers’ wages after considering the fixed contribution corresponding to the first pillar. In particular, we set it such that contributions from both pillars add up to the total contribution made under CSI (including the government subsidy). The result is a variable rate, \(r_{sj}^{CB}(w_{sj})\), obtained from the difference between the amount contributed for retirement pensions under CSI for each wage level \(j (j = 1, 2, \ldots, J)\) and the amount corresponding to a worker earning two minimum wages, expressed as a share of the wage (where the superscript CB denotes complementary SI benefits and \(w_{sj}\) the wage of salaried workers in wage level \(j\)). By construction, this complementary contribution would only be positive for workers earning more than two minimum wages, and would be increasing in wage levels expressed both as a share of the wage and in absolute terms (since two minimum wages represent a decreasing proportion of earnings as wages increase).\(^6\)

\(^5\) The government could induce non-salaried workers to save more for their retirement with matching contributions or incentives through the income tax law. But these savings would be voluntary as opposed to the forced savings imposed on salaried workers.

\(^6\) Since the second pillar is fully funded from ear-marked wage-based taxes, there are no implications for the government’s fiscal balance. In equation (28) one would just add on both the revenue and the expenditure side the term \(\sum_{j=1}^{J} r_{sj}^{CB}(w_{sj})w_{sj}L_{sj}\), which measures the contributions made to the retirement pensions of salaried workers under the second pillar.
The same procedure can be repeated for life and disability pensions. In this case, \( \tau_{sj}^{CB} \) is re-interpreted as the complementary contribution rate for both retirement, and life and disability pensions, and is calculated as described above. The only difference is that revenues for life and disability pensions are deposited in the IMSS-managed common reserve fund for risk pooling purposes, while revenues for retirement pensions are deposited in workers’ individual retirement accounts.

Consider, finally, risks that are specific to salaried workers, in particular, accidents suffered in the workplace because the firm fails to comply with safety standards. These risks are covered through work-risk insurance, and two points are relevant in this context. First, for moral hazard reasons contributions need to be firm-specific, based on firm’s safety records. Second, contributions need to be made for all salaried workers, including those earning less than two minimum wages, as work accidents are not insured by USI. In this context, we re-interpret \( \tau_{sj}^{CB} (w_{sj}) \) as the wage-based contribution rate that covers, for all salaried workers, work-risk pensions; and for those earning more than two minimum wages, complementary contributions for retirement and life and disability pensions.

Table 15 shows the implications of complementary pension benefits for workers’ utility and firms’ labor costs. Following the discussion above, we refer to workers as salaried and non-salaried rather than formal or informal, given that in a context where all workers benefit from USI the adjectives formal and informal are no longer relevant. Note that \( \beta_s^{CB} \in [0,1] \) is salaried worker’s valuation of the complementary pension benefits (assumed the same for all workers), and \( w_n \), the wage rate or, more precisely, equivalent earnings of non-salaried workers.

<table>
<thead>
<tr>
<th></th>
<th>Worker’s utility</th>
<th>Cost of labor</th>
<th>“Pure” labor tax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaried</td>
<td>( w_{sj} [1 + \beta_s^{CB} \tau_{sj}^{CB} (w_{sj})] + \beta^{USI} \tau^{USI} \tau_{sj}^{CB} (w_{sj}) ] )</td>
<td>( w_{sj} [1 + \tau_{sj}^{CB} (w_{sj})] )</td>
<td>( (1 - \beta_s^{CB}) \tau_{sj}^{CB} (w_{sj}) )</td>
</tr>
<tr>
<td>Non-salaried</td>
<td>( w_n + \beta^{USI} \tau^{USI} )</td>
<td>( w_n )</td>
<td>0</td>
</tr>
</tbody>
</table>

Clearly, if \( \beta_s^{CB} = 1 \), complementary pension contributions do not create a “pure” tax on salaried labor, nor distort firms and workers choices. Their incidence is fully shifted to workers, with no changes to the composition of employment vis-à-vis the USI equilibrium.
with \( \tau_{s,j}^{CB}(w_{s,j}) = 0 \) for all \( s,j \). However, if for reasons having to do with lack of trust in the system or hyperbolic discounting in the case of retirement pensions, or actuarially unfair fees for work-risk and life and disability pensions, \( \beta_x^{CB} < 1 \), a pure tax on salaried labor is introduced, with the accompanying distortions in firms and workers choices. That said, note that by construction \( \tau_{s,j}^{CB}(w_{s,j}) < \tau^{CSI} \); additionally, one expects that \( \beta^{CB} > \beta^{CSI} \), since in the bundle of complementary pension benefits there are no cross-subsidies for health, housing or day care. As a result, even if \( \beta_x^{CB} < 1 \), the pure tax is smaller than the one present under CSI and, in particular, lower for workers at the bottom end of the wage distribution.

Table 16 presents the values of \( \tau^{CSI} \) and \( \tau_{s,j}^{CB}(w_{s,j}) \) by wage level measured in multiples of the minimum wage where, to reiterate, \( \tau_{s,j}^{CB}(w_{s,j}) \) is computed such that contributions for work-risk, retirement and life and disability pensions are the same as what formal workers receive under CSI. In addition, the table shows the (cumulative) distribution of workers by formality status and wage level in 2008.

<table>
<thead>
<tr>
<th>Wage Level</th>
<th>Up to 2</th>
<th>2 to 3</th>
<th>3-4</th>
<th>4-5</th>
<th>9-10</th>
<th>14-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau^{CSI} )</td>
<td>0.320</td>
<td>0.320</td>
<td>0.320</td>
<td>0.320</td>
<td>0.320</td>
<td>0.320</td>
</tr>
<tr>
<td>( \tau_{s,j}^{CB}(w_{s,j}) )</td>
<td>0.015</td>
<td>0.045</td>
<td>0.061</td>
<td>0.069</td>
<td>0.087</td>
<td>0.093</td>
</tr>
<tr>
<td>Formal (%)**</td>
<td>34</td>
<td>56</td>
<td>69</td>
<td>77</td>
<td>91</td>
<td>96</td>
</tr>
<tr>
<td>Informal (%)**</td>
<td>49</td>
<td>72</td>
<td>86</td>
<td>90</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>Total (%)*</td>
<td>42</td>
<td>65</td>
<td>79</td>
<td>84</td>
<td>95</td>
<td>98</td>
</tr>
</tbody>
</table>

*Measured in multiples of the minimum wage; **Cumulative.

Note first that for workers who earn up to two minimum wages our proposal implies a more than twenty-fold reduction in wage-based taxes, as these taxes only cover work-risk insurance. Second, for workers earning up to four minimum wages, representing 79% of the labor force, the reduction in wage-based taxes is also very large, from 32% of the formal wage to 6.1%. But, third, even for workers earning up to fifteen minimum wages, representing 98% of the labor force, these taxes would fall by more than two thirds. Thus, our proposal allows for a very large reduction in labor taxes for all currently formal workers, from 32% to a
weighted average of 5.5%, who at the same time would receive at least the same health and pension benefits as under CSI.

The more relevant comparison, however, is between the pure taxes and subsidies in both cases. As noted, under the CSI + NCSI configuration the “pure” tax on salaried labor is 20%, and is accompanied by a subsidy to non-salaried labor of 7% (of the informal wage). To determine the “pure” tax under our proposal we need to estimate the value of $\beta^{CB}_s$, which is not feasible since this scenario has not been observed. That said, assume arbitrarily that this parameter equals 0.85 (the same value calculated for $\beta^{NCSI}_s$). In this case, the “pure” tax ranges from a low of 0.22% for workers earning up to two minimum wages, to a high of 1.4% for workers earning fifteen minimum wages; even if $\beta^{CB}_s = 0.5$, these taxes would range from 0.75% to 4.6%. In addition, there would be no subsidy to non-salaried labor. Differently put, even when complementary pension benefits for salaried workers paid through a Bismarkian wage-based tax are incorporated into our proposal, there is still a drastic change in the tax-cum-subsidies on salaried and non-salaried labor vis-à-vis the CSI + NCSI configuration.\(^{58}\)

Table 16 is of considerable importance from a macroeconomic point of view as well. In our model all goods are traded and we cannot properly speak of a real exchange rate. That said, we can compute the ratio of an index of good prices to an index of the costs of salaried labor (wages plus SI benefits), which could be thought of as a proxy measure of the real exchange rate. This ratio changes from 1.00 under the CSI + NCSI configuration, to 1.09 under USI. This would be a measure of the, so-to-speak, real depreciation implied by our proposal, interpreted here as a reduction in the price that firms pay for labor in terms of the exogenously given goods prices. In parallel, note that the VAT increase has a different impact on export vs. imports, as only the former are VAT-exempt. As a result, the relative price of

\[^{58}\text{An extension of our model to incorporate skill and wage differences and capture how } \tau^{CB}_{ij}(w_{ij}) \text{ impacts firms’ decisions to evade these contributions and the VAT is left for further research. As noted, the weighted average of } \tau^{CB}_{ij}(w_{ij}) \text{ is 5.5%. Incorporating this value into our model assuming that } \beta^{CB}_s = 0.85 \text{ yields VAT revenues that are just 1% lower than in the USI equilibrium without complementary contributions, so a preliminary result is that the distortions introduced by these complementary contributions are fairly small (the average “pure” tax on salaried labor is 0.8% vs. 20.0% under CSI).}\]
exports in term of imports would decrease, mimicking the result obtained in traditional trade models by combining a tax on imports and a subsidy to exports.⁵⁹

Table 16 also has substantive implications for individual worker’s retirement pensions, and for the level of aggregate savings associated with SI. On one hand, individual workers would enjoy higher retirement pensions vis-à-vis CSI because they would accumulate for their pensions throughout their working lives: when formally employed they would accumulate in their individual accounts the same amount as under CSI (although the expectation is that average time in salaried employment would increase); in addition, when non-salaried employed they would also accumulate through the first pillar. Differently put, contribution densities would be in principle 100% (vs. 45% under CSI), with at least the same amount of contributions as under CSI. Replacement rates would thus be unambiguously higher.⁶⁰

On the other hand, table 17 shows that the aggregate annual flow of savings for retirement under our proposal is higher than under CSI, for two reasons: one, an increase in the number of salaried workers under USI vs. formal workers under CSI (see table 12); two, the contributions made for non-salaried workers under USI. The result is an increase in savings for retirement of 166%, or almost an additional 1% of GDP⁶¹. Of course, forced savings for retirement could be offset by lower voluntary savings. But unless this offset occurs on a one-to-one basis, there would be a net increase in aggregate domestic savings.⁶² In turn, these additional savings could strengthen investment by firms since these savings would be long term resources deposited in the Afores.

---

⁵⁹ These results are related to the literature on “fiscal devaluations”, recently relevant in the context of attempts by some member countries of the Euro currency union to regain competitiveness lowering labor costs by shifting SI taxation from labor to consumption, given the impossibility of a nominal exchange rate devaluation; see Feldstein and Krugman (1990), Farhi et al. (2011) and IMF (2011).

⁶⁰ Incentives to firms to underreport salaried workers wages would also be lower, as wage-based taxes fall from 32% of the wage to an average of 5.5%; this would also augments resources accumulated for pensions.

⁶¹ Over the medium term the proposal would also deepen considerably the market for annuities and resources channeled to private insurance firms, as all 39 million workers would purchase an annuity upon retirement, as opposed to 12.3 million formal workers under CSI. Similar observations hold for annuities derived from life, disability and work-risk pensions.

⁶² Aguila (2011) studies the impact of Mexico’s 1997 pension reform and finds that only about 32% of forced savings in the new defined contribution system through individual retirement accounts was offset by reduced individual savings.
Table 17: Aggregate annual flow of savings for retirement*

<table>
<thead>
<tr>
<th></th>
<th>CSI</th>
<th>USI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Salaried workers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From wage-based contributions</td>
<td>54.9</td>
<td>56.6**</td>
</tr>
<tr>
<td>From government contributions</td>
<td>17.5</td>
<td>89.4</td>
</tr>
<tr>
<td><strong>Non-salaried workers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From wage-based contributions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>From government contributions</td>
<td>0</td>
<td>46.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>72.4</td>
<td>192.8</td>
</tr>
</tbody>
</table>

* thousands of million of pesos; **assuming the same wage distribution for all salaried workers under USI as the one observed under CSI.

We finish this section with a remark on risks faced by salaried workers as a result of negative output shocks. Mexico has no unemployment insurance. Rather, salaried workers are covered against this risk through severance pay and related regulations on firing. As discussed by Heckman and Pages (2004), these regulations represent a contingent cost to formal hiring that needs to be added to the explicit costs of CSI. Indeed, the fact that these regulations do not generate a counterpart monetary flow recorded as revenue somewhere (like health insurance or retirement pensions), should not detract from the more important fact that they increase the expected costs of salaried labor, affecting firms’ hiring decisions in much the same fashion that explicitly legislated CSI taxes do. It should also not detract from the fact that severance pay is an integral component of Mexico’s SI architecture.

The evidence for Mexico suggests that severance pay is an ineffective mechanism to protect salaried workers against the loss of employment (Kaplan et al., 2008a,b; Calderon, 2010). In this context, our proposal facilitates a reform of severance pay, protecting workers more effectively through unemployment insurance, for three reasons. First, because under the USI architecture, contributions for unemployment insurance can be added to \( \tau_{y}^{CR}(w_{y}) \) as another complementary SI benefit for salaried workers, and thus be paid by firms on a flow basis, perhaps through deposits in workers individual accounts, ensuring better protection for them.\(^63\) Second, because under our proposal, even if workers lose their salaried job, they are still covered by health, life and disability insurance and, at least for some period, continue to receive contributions for their retirement pensions (through the first pillar). And third, because salaried employment would be higher than formal employment under the CSI + NCSI

\(^{63}\) Firms in Mexico are not legally required to make reserves for severance payments. This creates difficulties to workers, as firms must make large lump-sum payments to workers precisely when they are facing negative output shocks and may be cash or credit constrained. The result is an incentive for firms to either litigate severance payments in labor courts when they have to fire workers or, in the case of smaller firms, to declare bankruptcy. In any event, the implication is that workers fail to receive income support precisely when they most need it.
configuration, implying that more workers would be protected against these risks (see table 12). Substituting severance pay, partly or wholly, with unemployment insurance in the context of USI is an important area of future research, with potentially large policy implications.

9. Some implications of USI for poverty and inequality, and for productivity

9.1 VAT, poverty and inequality

Perhaps the central argument against a VAT reform in Mexico, particularly one focused on raising the rate on food and medicines, is associated with its regressive effects on poor households. In our proposal this effect is in principle more than offset by the increase in the real wage resulting from lower CSI taxes (or, put differently, from a lower “VAT” at the door of the factory). Nevertheless, in a more realistic context where household composition differs, or where some households only have elderly members, the VAT increase could negatively impact the poor. To insure this is not the case, our proposal includes compensating them with direct income transfers to fully offset the effects of the VAT increase, ignoring the benefits of a higher real wage. To this effect, we calculate the additional VAT paid by households in the first two deciles of the income distribution, which we assume to be the population living under conditions of extreme poverty, offset these extra taxes by direct income transfers (carried out perhaps through a targeted program like Progresa-Oportunidades), and add the costs of these compensations to the fiscal costs of our proposal.

Data from Mexico’s 2008 income-expenditure survey indicates that households in the first two deciles of the income distribution account for 6.9% of total consumption. Our proposed VAT reform, on the other hand, extracts an additional 416.4 thousand million pesos (tmp) from all households, of which 28.8 tmp correspond to poor households (this is the amount added to the fiscal costs of the proposal in the last column of table 12). Setting compensations at this level, our proposal would have an unambiguously positive effect on the poor: they pay no additional net taxes; receive the same SI benefits as other households (including making effective the minimum pension guarantee); earn a higher real wage; and benefit from the largest reduction in CSI taxes, as these taxes are practically eliminated at the bottom of the wage distribution (see table 16).

There are three additional positive effects on poverty. The first is associated with the incentive structure faced by poor workers in the labor market. Levy (2008) argues that because
some NCSI programs are targeted on the poor, on one hand, and because the poor tend to value less the benefits of CSI programs than the non-poor, on the other, the tax on formality and the subsidy to informality is higher for them than for non-poor workers. As a result, poor workers are over-represented among informal workers: they are 24% of the labor force but account for 36% of informal employment (Levy, 2008). Because the marginal product of labor in informal employment is below that of formal employment, this implies that the CSI + NCSI configuration induces poor workers into lower productivity jobs. In this context, USI improves the conditions under which poor workers participate in the labor market, removing impediment derived from SI policy to their finding higher productivity jobs.

Second, our proposal would align the incentives of Progresa-Oportunidades with the incentives generated by SI policy. At present, the investments in the human capital of poor youngsters made by Progresa-Oportunidades are unlikely to yield the expected returns as when these youngsters enter the labor market they are, in the CSI + NCSI configuration, induced into informal jobs. Our proposal would contribute to increase the value of these investments, and through a better functioning labor market, raise the expected returns to education. In parallel, poverty policy could achieve a better pairing of instruments and objectives: USI would focus on protecting poor workers against risks, while Progresa-Oportunidades could focus exclusively on investments in the human capital of the poor (eliminating the use of this program to respond to shocks that should be covered by SI).64

The third effect has to do with vulnerability to negative shocks of non-poor Mexican households who have incomes just slightly over the poverty line. López-Calva and Ortiz-Juárez (2011a) show that uninsured shocks are a cause of downward mobility and high entry rates into poverty when these households experience negative systemic events like bad weather, but also idiosyncratic shocks like illness, death and disability. If these households could better insure these latter shocks—as they would under our proposal— their negative impact would be lessened or eliminated, resulting in reduced downward mobility and, overtime, a lower aggregate poverty rate.65

64 Our proposal implies therefore that Progresa-Oportunidades would continue as a targeted poverty alleviation program, in parallel to USI.

65 Further, using longitudinal data for 2002-2005, López-Calva and Ortiz-Juárez (2011b) calculate the probability that a non-poor unskilled urban informal worker in Mexico with primary education and access to insurance has of falling into poverty, and compare that probability to the one that would obtain if that worker lost access to insurance, or the one that would obtain if he kept access to insurance and in addition had three more years of
USI would also reduce inequality. Table 18 presents the results of a simple exercise where we distribute the additional burden of the VAT under our proposal over all households in proportion to their share in total consumption, direct compensations (only to the poor), and the change in subsidies for SI between CSI + NCSI and USI assuming benefits are distributed homogeneously across income deciles. Though over-simplified, the table serves to make two points. First, our proposal would have a strong redistributive effect: all households up to the first six deciles of the distribution would benefit, while most of the net burden would be paid by households in the highest two deciles. The reason is simple: in a context of high inequality in the distribution of income and consumption like Mexico’s, in absolute terms the VAT extracts substantially more resources from higher than from lower income households.\textsuperscript{66} If those resources are evenly distributed to provide the same SI benefits to all regardless of income level, it is not surprising that the net result is reduced inequality.

Table 18: Distribution of VAT burden and SI benefits by income deciles*

<table>
<thead>
<tr>
<th></th>
<th>I-II</th>
<th>II-IV</th>
<th>V-VI</th>
<th>VII-VIII</th>
<th>IX-X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional VAT</td>
<td>(-) 28.8</td>
<td>(-) 47.84</td>
<td>(-) 64.06</td>
<td>(-) 87.76</td>
<td>(-) 187.2</td>
</tr>
<tr>
<td>Compensations to the poor</td>
<td>(+) 28.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Net gain in SI benefits</td>
<td>(+) 66.4</td>
<td>(+) 66.4</td>
<td>(+) 6.4</td>
<td>(+) 66.4</td>
<td>(+) 66.4</td>
</tr>
<tr>
<td>Net impact</td>
<td>(+) 66.4</td>
<td>(+) 18.56</td>
<td>(+) 2.3</td>
<td>(-) 21.36</td>
<td>(-) 120.8</td>
</tr>
</tbody>
</table>

*Figures in thousands of millions of pesos.

The second point is related to the redistributive motivation of Bismarckian SI. Assume for a moment that all households in Mexico had the same income. Then our proposal would just change the composition of all households’ consumption equally, reducing their disposable income as a result of the higher VAT by the same amount, and giving them in return the same bundle of SI benefits contained in $\tau^{USI}$. This would achieve the government’s aim that all households be protected against risks, but without any redistribution of income between them. In principle, even in a society characterized by perfect equality in the distribution of income, SI education. Interestingly, they find that the effect of having insurance is quantitatively larger than acquiring more education to prevent workers from falling into poverty, highlighting the importance of extending the coverage of insurance to the near-poor.

\textsuperscript{66} This is why, regardless of considerations about social insurance, exemptions to the VAT are a very inefficient instrument to redistribute income in Mexico; see Davila and Levy (2003).
is still desirable to correct for market failures, aggregate risks, and so forth. This makes clear that the income redistribution that in Mexico’s case would be achieved under USI is a by-product of the underlying inequality in the distribution of income and consumption, rather than the express objective of SI. The point here is that in general one cannot expect one instrument, SI, to achieve two objectives at the same time – change the composition of all workers’ consumption and redistribute income from high to low income households. That said, the redistribution that would be observed in Mexico under USI would of course be welcome but, more importantly, substantially more effective than the one that is tepidly achieved today, if at all, under Bismarckian SI.67

9.2 Labor market distortions, USI and productivity

Lagging productivity is the main reason why Mexico’s growth performance has been lackluster over the last decades.68 Many factors account for this, and it is difficult to disentangle the relative weights of each. However, evidence suggests that misallocation of capital and labor resulting from distorted factor and output prices is a significant source of total factor productivity losses (Hsieh and Klenow, 2009b; IDB, 2010). In turn, distortions in the price of labor derive from Mexico’s dual SI architecture, as in the CSI + NCSI configuration this price varies by the type of labor contract and firm size; see table 5. The results are large differences in the marginal revenue product of labor across firms and sectors which misallocate factors and reduce productivity. Busso, et al. (2011) classify Mexican firms in the 2008 Economic Census by formality status, group them into six-digit sectors, and estimate indices of total factor productivity for each firm. They find that controlling for size and sectors, informal firms with non-salaried contractual relations are less productive than informal firms with illegal salaried labor, which in turn are less productive than formal firms with legal salaried labor. All

67 Of course, there would still be room for re-distribution through other taxes, in particular income taxes, and through other components of expenditure policy.

68 Over the period 1960-2007 factor accumulation was faster in Mexico than the United States (US); if total factor productivity (TFP) had kept pace, relative income per capita would be 24 per cent higher in 2007 vs. 1960. However, the sharp fall in Mexico’s TFP relative to the US experienced since 1980 more than offset the gains from factor accumulation, with the result that in 2007 Mexico’s relative income per capita was 14 per cent lower; see Busso et al. (2011).
in all, they find that formal firms produce about 35% more output from the same amount of capital and labor than informal firms.\footnote{Their exercise excludes self-employment and firms in non-fixed locations not captured in the Census which, as shown in table 2, account for a large share of employment. Incorporating these activities in the measurements of the productivity losses associated with the CSI + NCSI configuration is a pending challenge.}

Firm’s evasion strategies affect critical dimensions that impact productivity like size and legal status (with implications for firms’ access to credit from the formal financial system, for investments in innovation or technology adoption, and for links with suppliers of inputs); or the duration and type of labor contracts (with implications for firm’s investments in labor training or on-the-job learning). In general, an atmosphere of illegality and informality leads to inefficiently high levels of self-employment and non-salaried employment in family firms, excessive labor turnover, a size distribution of firms skewed towards smallness, and a bias towards sectors and activities where evasion and elusion of labor and CSI regulations can occur more easily.

Our proposal would have a positive effect on productivity by closing the wedge between the price of salaried and non-salaried labor; by removing the implicit tax on firm size derived from SI policy; and by eliminating labor turnover associated with firm’s evasion strategies. We speculate as well that a context of more legality and reduced evasion could reduce uncertainty, improve firm’s access to credit and expand their planning horizons, facilitating in turn more vertical integration, fuller exploitation of economies of scale, and increased investments in labor training, technology adoption and innovation. Quantifying these effects exceeds the scope of this paper, but the suggestion is clear: USI can simultaneously be better social policy, better fiscal policy, and better policy for productivity and growth.

10. Implementation issues

This section discusses implementation issues, with emphasis on mechanisms to identify the beneficiary population of USI and provide benefits. Many other aspects are left out, as well a discussion of transition issues, which merit a separate paper. The discussion here is indicative and meant to show the administrative feasibility of the proposal. We stress that there is no unique implementation mechanism. Many factors that bear on implementation design are neither budgetary nor technical but political (or philosophical), having to do with the appropriate balance between the exercise of social rights and the enforcement of obligations.
Although inevitably technical and political issues overlap, we do not dwell into that discussion and simply point out some of the relevant trade-offs.

10.1 Registration and reporting requirements

We begin with a key aspect of our proposal. To receive the benefits of USI, workers have to be registered with the Ministry of Finance and have an individual retirement account with the Afore of their choice. Under ideal circumstances there would be a one-to-one mapping through a unique identification number between individual workers, registries of workers in the Finance Ministry, registries of workers in IMSS (including non-salaried), and workers’ Afore accounts.

For the purposes of our proposal, a worker is a person that is either registered as such with the Finance Ministry by the firm that hires him (for the case of salaried workers), or by himself in case he is self-employed, works for a small family firm or is associated with a firm but receiving payments in forms other than wages; in these latter cases, corresponding to non-salaried contractual relations, the obligation to register would fall on the worker. All registered workers, in turn, have to report their earnings. For salaried workers this requirement, as at present, would be satisfied by the firms that hire them, since they already report to retain income taxes from gross wages. Non-salaried workers would have to report their income directly to the Ministry of Finance on a periodic basis, say quarterly.\(^{70}\)

Importantly, the act of registration and reporting would be sufficient to qualify for USI benefits; failure to pay income taxes would not disqualify workers. Our proposal is not designed to solve the problems of compliance with the Income Tax Law. But it would contribute to improve its enforcement as relative to the status quo it enhances the incentives to register and report. On one hand, as discussed in section 8, since firms hiring salaried workers would continue to pay wage taxes for complementary SI benefits, the registries of IMSS could be used by the Finance Ministry for enforcement of income taxes. In parallel, firms’ incentives to cheat would be substantially reduced, as wage-based taxes fall from 32 to an average of 5.5% of wages.

\(^{70}\) Of course, workers who register could under-report earnings. But note that this problem is already present today and matters from the point of view of income tax collections. Under-reporting of earnings is not a problem from the perspective of financing USI benefits since under our proposal this occurs through the higher VAT on workers consumption (which presumably is a better proxy for their true earnings than those reported to the Finance Ministry). In fact, workers could register and cheat but still receive USI benefits, although if that behavior was recurrent it should sooner or later trigger an audit by the Finance Ministry.
For non-salaried workers, on the other hand, the situation is more complex since the Income Tax Law taxes income differently depending on its source. These workers are already required to register but many decide not to. Our proposal changes the cost-benefit of that decision in favor of registration, as tax obligations in principle do not change while, depending on the choices described below, there would be no or restricted access to SI benefits.71

The fact that more workers than at present register and report earnings to the Finance Ministry is a positive side effect of our proposal, generating valuable information for tax enforcement (and for the design of public policies in general). It would be a step in the direction of reducing illegality and promoting compliance with the Law. But, equally important at least in the view of some, complying with an obligation in order to receive a benefit would balance workers’ rights with responsibilities and reduce “free-riding” behavior.72

10.2 Pension benefits

For salaried workers these benefits would operate as they currently do for formal workers, except that wage-based contributions for retirement pensions, and for life and disability insurance, would be reduced, with the concomitant increase in contributions by the government, along the lines of the discussion in section 8 (though contributions for work-risk insurance would not change). Similar procedures would operate for non-salaried workers (excluding work-risk insurance). Upon compliance of the registration and reporting requirement, the Ministry of Finance would transfer to IMSS the contributions for retirement pensions and life and disability insurance, in much the same way as it does today for formal workers. In turn, IMSS would use the same mechanisms already in place to transfer the

71 There would be transaction costs of registering (only once) and reporting (four times a year). But through the use of information technologies, these costs could be minimized. On the other hand, Paullada (2012) argues that the Income Tax Law taxes unequally earnings from wages relative to other sources of income. This important point suggests that incentives to arbitrage between salaried and non-salaried status and change labor contracts would be much reduced if, in parallel to our proposal, the Income Tax Law was revised in the direction of more neutral treatment between income of salaried and non-salaried workers. Equally important would be a consolidation of the various regimes for non-salaried workers.

72 There is an important precedent in that context in Mexico, associated with the right to vote. Before this right can be exercised, citizens need to register with the Federal Electoral Institute, the relevant authority for these purposes. Mutatis mutandis, before the right to USI benefits can be exercised, workers would need to register with the Finance Ministry.
contribution for retirement pensions to workers’ individual retirement accounts, and deposit in the IMSS-managed common reserve fund the contribution for life and disability insurance.\textsuperscript{73}

Nevertheless, some non-salaried workers could prefer not to register and avoid any reporting requirements. Those workers would in principle be excluded from the pension benefits of USI, including the minimum pension guarantee (and maybe health benefits, see below). But, critically, note that this would result from a personal choice, triggered by distrust of government, procrastination or other motivation. It would be an individual decision not to exercise the right to receive a contribution for a pension, not an exclusion from that right by the design of the SI system.\textsuperscript{74}

The decision of workers to register and report is not independent of their expectations of whether, when they reach old age, they would have access to any other type of pension, without having registered at all and accumulated for a USI retirement pension. The discussion so far has implied that all NCSI benefits would be replaced by USI benefits. But an exception needs to be made for the existing NCSI retirement pension, at least for some time. This is because even if our proposal was implemented immediately, low wage older workers would have relatively little time to accumulate sufficient resources in their individual retirement accounts before they reach retirement age to purchase an annuity that is at least as good as the existing NCSI pension, even if they register and accumulate for a USI pension. Further, many already retired workers have no pension. As a result, for a transition period it would likely be necessary to continue with some form of a NCSI pension (perhaps means-tested); indeed, even to expand its coverage.\textsuperscript{75}

\textsuperscript{73} Informal workers would need to have an individual retirement account with the Afore of their choice. But note that as a result of large transits between formal and informal employment, many currently informal workers already have an account, given previous episodes of formality. In 2006 there were 37.4 million Afore accounts, for a total labor force of about 40 million workers (excluding public sector workers); see Levy (2008). Some of these accounts are duplicated, but this is estimated to be in the order of 10%. The implication is that probably over two thirds of all workers already have an account.

\textsuperscript{74} One could argue that poor workers in remote rural areas could have difficulties exercising their rights given the complexities that for them represent the registration and reporting requirements. Note, however, that at present these workers already receive the benefits of Progresa-Oportunidades, which imply that mechanisms exist to continuously up-date records to make the associated by-monthly cash payments. Registering households in remote rural areas, and monitoring indicators on their behavior, is a problem that other institutions of the Mexican State have already solved. Moreover, poor workers would have strong incentives to register and report, as compliance would imply no tax payments (since their low earnings would exempt them from income taxes) but better health and pension benefits, including the possibility of making effective the minimum pension guarantee.

\textsuperscript{75} The main NCSI pension program at the federal level is Adults over Seventy, although many state governments run similar programs. This program targets households by location, focusing on small rural communities, and
This issue is complex and merits more analysis. Should NCSI pensions be gradually phased-out, or should a “low” old-age means-tested pension paid on a pay-as-you-go basis permanently co-exist with pensions financed from savings accumulated in workers’ individual retirement accounts, given that under USI all workers are entitled to receive a contribution for their pension? The point is that these two types of pension should be seen as part of the incentives faced by workers, particularly in a context where two potentially conflicting objectives need to be met: on one hand, a major effort to induce all workers to accumulate for their pensions through their individual retirement accounts, particularly young ones; on the other, the need to provide income support to avoid old-age poverty for retired, or soon-to-be-retired workers. Can an inter-temporally consistent mechanism be designed to provide income support to already retired workers without a pension (and to informal workers close to retirement), while at the same time inducing younger workers to accumulate for their pension in their individual retirement account, thus avoiding moral hazard behavior?

10.3 Health benefits

At present, the health benefits of CSI cover the worker and his family, and retired workers with a CSI pension. Since under our proposal CSI health benefits extend to all workers, retired workers without a CSI pension as well as household members of currently informal workers would also be covered by the health component of USI. In fact, the per worker cost of USI included in our previous calculations, based as it is on the per worker cost of the health component of CSI, already contemplate this. Thus, our funding proposal, for all intents and purposes, would cover the whole population.

---

Pay a monthly pension of 500 pesos, or one third the minimum CSI pension (1,500 pesos a month). But, as noted, at present most low wage workers will not receive the minimum CSI pension since as a result of their high transit between formal and informal status they will not accumulate the 25 years in formal employment needed to qualify; for them the NCSI pension is a better alternative. The same would hold for older low wage workers under our proposal, but not for young ones for whom the USI pension would be significantly better (at least equal to 1,500 pesos a month but probably higher). Maintaining the Adults over Seventy Program for some time would have minor budgetary implications. On one hand, table A2.2 shows that in 2008 the cost of this program was 9.5 tmp (or 0.07% of GDP), which would need to be added to the costs of our proposal as quantified in table 12. On the other, table 12 assumes that the contribution for the USI retirement pension is made for all workers, including those that are close to retirement, which would be unnecessary if the Adults over Seventy program is maintained for a transition period.

76 Galiani and Gertler (2009) is a step in this direction. They show that the introduction of the NCSI pension program “Adults over Seventy” led to lower participation rates in salaried employment for older workers close to retirement age (through an “anticipation effect”).
Regardless of budgetary considerations, a difficult question is whether unregistered workers should be excluded from the health benefits of USI. At present, informal workers benefiting from NCSI health programs are not subject to any registration or reporting requirements, as opposed to formal workers who need to be registered with IMSS and the Finance Ministry to receive CSI health benefits. USI health benefits could then be conceptualized as either extending the NCSI modality, and therefore open to registered and unregistered workers; or extending the CSI modality, and therefore denying access to unregistered workers. Under the CSI modality the incentives to register and report are substantially strengthened, but there is an issue as to whether the government can credibly (and fairly) deny health benefits for workers that do not register. Under the NCSI modality access to health is greatly facilitated, but the incentives to register and report weakened as the only benefits of doing so are a retirement pension (in the future) or a disability or survivorship pension (at some undetermined time). Our proposal is compatible with either modality (or intermediate situations), and the choice requires further discussion and debate.

Under any modality, the provision of health services would be carried out by IMSS and state governments, as occurs at present under the CSI + NCSI configuration. All that is needed is the number of salaried and non-salaried workers in each state. On that basis, the Ministry of Finance would provide to IMSS and to each state government the corresponding per worker fee. Note that if the CSI modality is adopted the incentives for state governments to cooperate with the Federal government and have workers in their state register with the Finance Ministry are stronger, as those registries would be the basis for resources transfers.

IMSS and state government would be responsible for providing services, with the Health Ministry responsible for regulatory issues and for monitoring quality of provision; since all state government and IMSS receive the same resources per worker, the expectation would be that quality should converge to the same national level, gradually reducing existing regional

---

77 Recall, however, the public sector workers are excluded from our proposal, as they are covered by their own social security institute.

78 Because CSI health services already receive a subsidy from the Federal Government, the mechanism for transferring resources from the Ministry of Finance to the IMSS on a per worker basis is already in place. On the other hand, at present the Ministry of Finance transfers resources for some health programs to the Health Ministry, which in turn transfers them to state governments. These resources are complemented by state resources. Other NCSI health programs, however, are directly run by the Federal Government (including, confusingly, one by IMSS). The proposal thus requires unifying these budgetary mechanisms in a single one. This would ensure the same level of per worker financing as that provided to IMSS and enhance transparency.
differences. There would be no changes to the ownership structure of public health infrastructure, but state governments and IMSS would be given the freedom to write service contracts between them to mutually complement their infrastructure, and as needed to sub-contract activities or purchase inputs from private providers to improve quality.\textsuperscript{79}

To conclude, the discussion in this section is not intended to assert that implementing USI will be trouble-free, or that there are no difficult trade-offs between competing objectives. On the contrary, it will require important adjustments to the functioning of various ministries and institutions, to budgetary relationships between the Federal and state governments, and to many laws. It will also require a political discussion about the appropriate balance between the exercise of social rights and the enforcement of social obligations. After all, transiting from the current CSI + NCSI configuration to USI would be a major social and institutional change. Moreover, as implementation occurs unforeseen obstacles will be encountered; many adjustments will certainly have to be made. Thus, if the reform was to be implemented, it would need—as any major reform would—careful evaluation and error-correcting mechanisms; much will be learnt along the way. That said, the discussion does mean to argue a simple but central point: implementing USI is feasible and within the reach of the administrative and operational capabilities of the Mexican State.

11. Concluding thoughts

“The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist…I am sure the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas. Not, indeed, immediately, but after a certain interval…. But soon or late, it is ideas, not vested interests, which are dangerous for good or evil”.

When President Avila Camacho signed Mexico’s first Social Security Law in 1943, he was probably unaware that seven years earlier Keynes had concluded The General Theory with the words quoted above. Back then he probably thought that Bismarck’s idea to fund SI from

\textsuperscript{79} A challenge in this context is setting incentives to IMSS and state governments to choose the efficient combination of in-house and out-source provisioning. Because per worker transfers to each would be established by Law, both providers could write long term contracts between themselves and coordinate investment decisions to ensure, on one hand, the availability of appropriate infrastructure in remote rural areas and, on the other, duplication of equipment and infrastructure in larger urban areas (as at times currently happens).
wage-based taxes would allow him to reach his goal of protecting all Mexican workers from various risks, to combat social exclusion, and to redistribute income from “capital” to “labor”.

Were he alive today, he would most likely be intensely disappointed realizing that almost seventy years after his Law came into effect, only one third of Mexican workers were covered by its provisions (despite important revisions in 1973 and 1997). He would also probably be distressed if he realized that in 2011 Mexico’s SI architecture, rather than strengthening social inclusion, was the root cause of the segmentation of workers into formal and informal categories, with social rights and obligations determined by the form of their labor contract; certainly, creating first and second class workers was not his intention. Moreover, he would likely be intensely concerned that this architecture generated large economic inefficiencies, as he was no doubt aware that sustained prosperity could not be built on stagnant productivity. In parallel, although undoubtedly he would agree with the social need to provide SI benefits to informal workers, and perhaps sympathize with the political pressures, he nonetheless would probably be very preoccupied if he realized that the continuous growth of NCSI programs obscured the relationship between contributions and benefits, gradually eroding the long term fiscal sustainability of Mexico’s SI policy; and in parallel undermined the Rule of Law. In sum, he would in all likelihood be strongly dissatisfied with the status quo, and would be eager for a new course.

We do not argue that in 1890 Bismarck’s idea was flawed. Quite the contrary, back then it was novel, bold and, for Germany, effective. We do not argue either that in 1943 President Avila Camacho was wrong importing Bismarck’s idea to Mexico; in fact, at that time it was probably the best that could be achieved given the prevailing knowledge and administrative capabilities of the Mexican State. But every idea has a time and a place. With the benefit of hindsight, we do argue that Bismarck’s idea has not served Mexico well. Every Mexican president from Avila Camacho then to Calderon today has struggled with Bismarck’s inheritance, attempting through various means to extend SI coverage to informal workers. The result is the CSI + NCSI configuration: costly in terms of productivity; fiscally unsound; socially ineffective.

The idea that social insurance should be funded from a wage-based tax is deeply ingrained in Mexico. This idea has generated a large body of legal thought and jurisprudence; a long-standing conviction that, somehow, social justice is achieved by taxing salaried labor; and some of Mexico’s key social institutions. In turn, vested interests have evolved and profited
from this whole construct. But, along with Keynes, we are convinced that “the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas”. Without minimizing the political obstacles that vested interest represent, at this point we are convinced that the encroachment of Bismarck’s idea is a central obstacle to a more prosperous and equitable Mexico. But Bismarck’s idea is just that, an idea; it is not a Law of Nature. Paraphrasing Keynes’s admonition about practical men being prisoners of the ideas of some defunct economist, we posit that in matters of social insurance the opposite may be true: economists are the prisoners of the idea of a long defunct politician. Mexico needs to break away from Bismarck’s idea. This paper shows that it is fiscally and administratively feasible to do so, and that there would be large gains from doing so. Transforming Mexico’s social insurance architecture will require political leadership as bold as the one displayed by President Avila Camacho back in 1943. But before that leadership can be displayed again by another Mexican president, it is first necessary to turn Bismarck on his head.
Appendix 1. Model structure

A.1 The intermediate goods sector

Two intermediate goods $I_z$, indexed by $z = 1, 2$, are produced by a large number of firms that behave in a competitive fashion. Firms sell their output to producers of final good $A$ at the exogenous price $p_z$. Goods are produced with a Cobb-Douglas technology:

$$I_z = A_z L_z^a K_z^{1-a}, \tag{A1.1}$$

where $\alpha$ satisfies $0 < \alpha < 1$. Physical capital $K_z$ is fixed, so the representative firm makes positive profits in equilibrium. $K_z$ is continuous and distributed exogenously among firms with a density function $f(K_z)$. Capital endowment in the economy is given by $\bar{K}$ and satisfies the resource constraint given by equation (2) in the paper. Labor input $L_z$ is the sum of formally and informally hired workers following equation (3) in the paper. If a firm hires a worker formally the unit cost is given by equation (4). If the firm hires the worker informally it faces an endogenous probability $\lambda^C_z(L_{iz}, K_z)$ of being detected by the authority and the expected unit cost of labor is given by equation (5). This probability is given by the function:

$$\lambda^C_z(L_{iz}, K_z) = \hat{\lambda}^C_z K_z^v L_{iz}^2, \tag{A1.2}$$

where $\hat{\lambda}^C > 0$ is a parameter, and $v > 0$. Note that (A1.2) implies that the probability of detection is increasing in both arguments.

Firms pay income and value-added taxes and have incentives to evade. They may be detected by the authority with a probability $\lambda^VAT_z(K_z) = \lambda^IT_z(K_z) = \lambda(K_z)$ given by:

$$\lambda(K_z) = \hat{\lambda}K_z / \bar{K}_z, \tag{A1.3}$$

where $\hat{\lambda} > 1$ is a parameter. This specification implies that relatively large firms (that is, firms with a capital $K_z \geq \bar{K}_z / \hat{\lambda}$) face a probability of detection equal to one. The effective VAT rate for each firm is defined by equation (6) in the text. The firm rate of compliance with income taxes, $\xi^IT_z(K_z) \in [0, 1]$, is:

$$\xi^IT_z(K_z) = \min[a^IT_z \eta \bar{z}(K_z) \eta(K_z), \tau^IT_z \eta(K_z)] / \tau^IT_z \eta(K_z). \tag{A1.4}$$

Accordingly, the effective tax rates faced for a firm of size $K_z$ are $\tau^VAT_z \xi^VAT_z(K_z)$ and $\tau^IT_z \xi^IT_z(K_z)$.\(^{80}\)

The problem of a representative firm in the intermediate goods sector is to choose formal and informal labor, $\{L_{fz}, L_{iz}\}$, to maximize expected profits:

\(^{80}\) Expressions (6) and (A1.4) imply that there is a level of capital $K_z^+$ at which the rates of compliance are one, that is $\xi^VAT_z(K_z^+) = \xi^IT_z(K_z^+) = 1$. Given the specification for $\lambda(K_z)$, this condition implies $K_z^+ = \bar{K}_z / \hat{\lambda}$ and $\xi^IT_z(K_z^+) = \bar{K}_z / \hat{\lambda}$. For firms with a relatively large capital endowment ($K_z \geq K_z^+$), their rates of compliance are equal to one so their corresponding effective tax rates are $\tau^VAT_z$ and $\tau^IT_z$. These firms fully comply with VAT and income taxes even though they have incentives to evade their payment.
subject to $L_z(K_z) = L_{fz}(K_z) + L_{iz}(K_z)$ given prices $\{p_z, w_f, w_i\}$ and taxes $\{\tau_{CSi}, \tau_{VAT}, \tau_{IT}\}$. From first-order conditions, the expressions for total labor demand $L_z(K_z)$ and informal labor $L_{iz}(K_z)$ are:

**(A.6)**

$$L_z(K_z) = \left( \frac{(1 - \tau_{VAT}^{\text{IT}})^{(1 - \alpha)}}{[1 + (1 - \alpha) \tau_{CSi}^{\text{IT}}]w_f} \right)^{1/(1 - \alpha)} K_z, \quad K_z \in [K_z, \overline{K}_z], \text{ and}$$

**(A.7)**

$$L_{iz}(K_z) = \left( \left[ \frac{[1 + (1 - \alpha) \tau_{CSi}^{\text{IT}}]w_f - w_i}{3\phi\lambda_{CSi}^{\text{IT}}} \left( \frac{1}{K_z^{0.5}} \right)^{1/2} \right] \right)^{1/2},$$

Formal labor demand is given by the difference $L_{fz}(K_z) = L_z(K_z) - L_{iz}(K_z)$. These functional forms imply that the share of formal workers in total workers increases with capital size.

Expression (A.6) indicates that the demand for salaried labor depends negatively on the effective tax rate $\tau_{VAT}^{\text{IT}}$ and $K_z$ and CSI taxes $\tau_{CSi}^{\text{IT}}$. This last result has important implications for the tax base and government revenue: an increase in $\tau_{CSi}^{\text{IT}}$ lowers labor demand and thus value-added for a firm of capital size $K_z$. It also decreases profits for firms with a relatively large capital size $K_z$, given the high probability of being detected by the authority evading labor taxes. These two effects erode the tax base for both value-added and income taxes. Given that formal labor is now more costly than informal labor, informality increases.

The aggregate rate of compliance for the VAT is defined in equation (10) in the text. For income taxes, let $\hat{\Pi}_{z}(K_z)$ denote gross profits for a fully compliant firm. The aggregate rate of compliance for income taxes in sector $z$, $\hat{\xi}_{z}^{\text{IT}} \in [0,1]$, is:

**(A.8)**

$$\hat{\xi}_{z}^{\text{IT}} = \frac{\int_{K_z}^{\overline{K}_z} \hat{\xi}_{z}^{\text{IT}}(K_z)\Pi(K_z)f(K_z)dK_z}{\int_{K_z}^{\overline{K}_z} \hat{\xi}_{z}(K_z)f(K_z)dK_z}$$

The aggregate rates $\xi_{z}^{VAT}$ and $\xi_{z}^{\text{IT}}$ are used to compute the rates of compliance in the final good $A$ sector and to determine “VAT included” prices of intermediate goods in equation (11) in the text.

---

81 Notice from (A.6) and (A.7) that $L_{iz}(K_z) > L_z(K_z)$ for relatively small values of $K_z$, implying a negative labor demand for formal workers. To avoid this scenario, define $K_{z}^-$ as the level of capital that uniquely solves $L_z(K_z) = L_{iz}(K_z)$. This implies that firms with a capital level $[K_z, K_{z}^-]$ will hire informal workers only, according to (A.6). In contrast, firms with a capital level $(K_{z}^-, \overline{K}_z]$ will demand a mix of formal and informal workers whose total amount is also given by (A.6). In such a case, informal labor is determined by (A.7) whereas formal labor is given by $L_z(K_z) - L_{iz}(K_z)$. 

78
A.2 The final good A sector

This sector is composed of a large number of representative firms that behave in a competitive fashion. Firms use intermediate goods $I_1$ and $I_2$ in combination with a fixed factor $A_m$ to produce $A$. The production function is Cobb-Douglas:

$$A = [I(I_1, I_2)]^{\alpha_m A_m^{1-\alpha_m}},$$

where $0 < \alpha_m < 1$. In turn, the function $I(I_1, I_2)$ is CES:

$$I(I_1, I_2) = [\gamma(I_1)^\mu + (1 - \gamma)(I_2)^\mu]^{1/\mu},$$

with restrictions $-\infty \leq \mu \leq 1$ and $0 \leq \gamma \leq 1$. The parameter $\gamma$ represents the weight of intermediate good $I_1$ in the production of $I$. The elasticity of substitution between intermediate goods $I_1$ and $I_2$ is given by $1/(\mu - 1)$.

Firms pay income and value-added taxes. The VAT rate in this sector is a weighted average of the tax rates faced by intermediate goods firms, so $\tau^{\text{VAT}}_A \equiv \gamma \tau^{\text{VAT}}_1 + (1 - \gamma) \tau^{\text{VAT}}_2$. On the other hand, the income tax rate is the same as in the intermediate goods sector. Final good A firms also have incentives to evade taxes. Let $\xi^{\text{VAT}}_A$ and $\xi^{\text{IT}}_A$ be the respective rates of compliance, which are given by the weighted average of the corresponding aggregate rates of compliance in the intermediate goods sector. This implies $\xi^{\text{VAT}}_A \equiv \gamma \xi^{\text{VAT}}_1 + (1 - \gamma) \xi^{\text{VAT}}_2$ and $\xi^{\text{IT}}_A \equiv \gamma \xi^{\text{IT}}_1 + (1 - \gamma) \xi^{\text{IT}}_2$. Accordingly, the tax rates effectively paid by firms in the final good A sector are $\tau^{\text{VAT}}_A \xi^{\text{VAT}}_A$ and $\tau^{\text{IT}}_A \xi^{\text{IT}}_A$. Since value added taxes are collected by the credit method, the firm receives a tax credit for the amount $\tau^{\text{VAT}}_2 \xi^{\text{VAT}}_2 p_2 I_2$.

The problem of a representative firm is to choose intermediate goods $\{I_1, I_2\}$ to maximize expected profits:

$$\text{Max}_{I_1, I_2} \{ (1 - \tau^{\text{IT}}_A \xi^{\text{IT}}_A) \{ (1 - \tau^{\text{VAT}}_A \xi^{\text{VAT}}_A) p_A[I(I_1, I_2)]^{\alpha_m A_m^{1-\alpha_m}} \} - \Sigma_z (1 - \tau^{\text{VAT}}_z \xi^{\text{VAT}}_z) p_z I_z \},$$

taking prices $\{p_1, p_2, p_A\}$, taxes $\{\tau^{\text{VAT}}_1, \tau^{\text{VAT}}_2, \tau^{\text{IT}}\}$, and rates of compliance $\{\xi^{\text{VAT}}_1, \xi^{\text{VAT}}_2, \xi^{\text{IT}}_1, \xi^{\text{IT}}_2\}$ as given. From first-order conditions, the relative demand of intermediate goods is:

$$\frac{I_1}{I_2} = \left[ \frac{\gamma (1 - \tau^{\text{VAT}}_1 \xi^{\text{VAT}}_1) (p_1 + \tau^{\text{VAT}}_1 \xi^{\text{VAT}}_1)}{(1 - \gamma) (1 - \tau^{\text{VAT}}_2 \xi^{\text{VAT}}_2) (p_2 + \tau^{\text{VAT}}_2 \xi^{\text{VAT}}_2)} \right]^{1/(1-\mu)}.$$

which is expression (13) in the main text.

A.3 The final good B sector

Family firms or self-employed workers only require labor $L_B$ to produce goods, which are sold at the exogenous price $p_B$. The production function has decreasing returns to scale in labor:

$$B = A_B L_B^\alpha,$$
Unit labor cost is just $w_i$ since workers in family firms or that are self-employed are not obligated to contribute to CSI. Given that this sector does not pay any of the three taxes, the profit function is $\max \Pi_B = p_B \alpha_B L_B \alpha - w_i L_B$. Hence, family firms or self-employed workers choose the quantity of labor $L_B$ that maximizes profits, taking prices $\{p_B, w_i\}$ as given. Accordingly, optimal labor demand is:

$$(A.14) \quad L_B = \left(\frac{a p_B \alpha_B}{w_i}\right)^{1/(1-\alpha)}.$$  

### A.4 Social insurance, wage rates and the labor market

Total demand for labor is obtained aggregating demand for salaried labor (A1.6) over all firms in the intermediate sectors plus demand for non-salaried from the final sector B, given by (A1.14). Given an inelastic labor supply $\bar{L}$, the aggregate labor constraint is:

$$(A.15) \quad \sum_z \left\{ \int_{K_z}^{\bar{K}_z} \left[ 1 - \tau_z^{VAT} \xi_z^{VAT}(K_z) \right]^{1-\alpha} K_z f(K_z) dK_z \right\} \left(\frac{a p_B \alpha_B}{w_i}\right)^{1/(1-\alpha)} = \bar{L}.$$  

The relationship between the formal and the informal wage is captured in expression (19) in the text. With this we solve (A.15) and (19) simultaneously for equilibrium wages $\{w^*_f, w^*_i\}$, given value-added, income, and CSI and NCSI taxes and subsidies.

### A.5 The government’s budget constraint

The government’s fiscal balance is:

$$(A.16) \quad FB = (R^{VAT} + R^{IT} + R^{CSI} + \bar{R}) - (G^{CSI} + G^{NCSI} + \bar{G}).$$  

Revenue from value-added taxes is collected from the intermediate and final good A sectors, denoted respectively by $R_i^{VAT}$ and $R_A^{VAT}$, so that $R^{VAT} = R_i^{VAT} + R_A^{VAT}$. Aggregating over all firms and sectors producing intermediate goods, revenues are:

$$(A.17) \quad R_i^{VAT} = \sum_z \left\{ \tau_z^{VAT} A_z p_z \int_{K_z}^{\bar{K}_z} \xi_z^{VAT}(K_z) [L_z(K_z)]^{\alpha} K_z^{1-\alpha} f(K_z) dK_z \right\},$$  

where demand for salaried labor $L_z(K_z)$ is given by (A1.6). In the final good A sector, VAT revenue is:

$$(A.18) \quad R_A^{VAT} = \tau_A^{VAT} \xi_A^{VAT} [I(I_1, I_2)]^{a_m} A_m^{1-a_m} - \sum_z \tau_z^{VAT} \xi_z^{VAT} p_z I_z,$$

where $I(I_1, I_2)$ is defined by (A1.10). Notice that expression (A1.18) subtracts the amount of tax credit $\tau_A^{VAT} \xi_A^{VAT} p_z I_z$ to each sector.

Consider now revenues from income taxes. As before, government collects revenue from both intermediate and final good A sectors, denoted respectively by $R_i^{IT}$ and $R_A^{IT}$. Hence $R^{IT} = R_i^{IT} + R_A^{IT}$. Aggregating over all firms and sectors, we get:
In the final sector government revenue from income taxes is:

\[
R^IT = \tau^IT \sum_z \int_{K_z} R^IT (K_z) \frac{\xi^IT}{\xi_z} (K_z) f(K_z) dK_z
\]

CSI taxes paid by firms and formal workers are collected in the intermediate goods sector only, as firms in the final good A sector do not require labor. Hence:

\[
R^{CSI} = (1 - \theta)\tau^{CSI} \sum_z \int_{K_z} \frac{\xi_z}{\xi} L_fz(K_z) f(K_z) dK_z,
\]

where \(L_fz(K_z)\) is the demand for formal workers.

Total spending on CSI is split in two parts:

\[
G^{CSI} = (1 - \theta)\tau^{CSI} \sum_z \int_{K_z} \frac{\xi_z}{\xi} L_fz(K_z) f(K_z) dK_z + \theta \tau^{CSI} \sum_z \int_{K_z} \frac{\xi_z}{\xi} L_fz(K_z) f(K_z) dK_z.
\]

Considering (A1.16) and (A1.21), it follows that only subsidies to CSI constitute a net pressure on the fiscal balance.

Government spending on NCSI, \(G^{NCSI}\), is given by the subsidy \(\tau^{NCSI}\) per worker times the total number of informal workers. Aggregating over all firms and sectors leads to:

\[
G^{NCSI} = \tau^{NCSI} \sum_z \int_{K_z} \frac{\xi_z}{\xi} L_i(K_z) f(K_z) dK_z + \tau^{NCSI} \left( \frac{\alpha p_A A_B}{\bar{w}_1} \right)^{1/(1 - \alpha)}.
\]

### A.6 GDP and prices

GDP is the sum of value added in each sector:

\[
GDP = \sum_p p_z A_z \int_{K_z} \frac{\xi_z}{\xi} [L_z(K_z)]^\alpha K_z^{1 - \alpha} f(K_z) dK_z
\]

\[
+ \left[ p_A(I_1, I_2) \right]^\alpha A_m^{1 - \alpha} - p_1 I_1 - p_2 I_2 + p_B A_B L_B^\alpha,
\]

where the first term represents value added from intermediate sectors, and the second and third terms value added in the final good A and B sectors, respectively.

To compute the “VAT included” price of good A, consider the following cost minimization problem for a representative consumer demanding intermediate goods \(I_1\) and \(I_2\):

\[
\min_{(I_1, I_2)} \left[ p_1 (1 + \tau_1^{VAT} \xi_1^{VAT}) I_1 + p_2 (1 + \tau_2^{VAT} \xi_2^{VAT}) I_2 \right]
\]

subject to \([\gamma(I_1)^\mu + (1 - \gamma)(I_2)^\mu]^{1/\mu} = \bar{I}\). Recall that prices \(p_1\) and \(p_2\) are given as the economy is small in world markets. Let \(\bar{p}_z \equiv (1 + \tau_z^{VAT} \xi_z^{VAT})p_z\) denote the “VAT included”
price of intermediate good \( z \). The demand functions arising from the cost-minimization problem are:

\[
\begin{align*}
\tilde{I}_1 &= \left[ \frac{\bar{p}_1}{\bar{p}_1} \right]^{\frac{1}{1-\mu}} I, \quad \text{and} \\
\tilde{I}_2 &= \left[ \frac{(1-\gamma)\kappa}{\bar{p}_2} \right]^{\frac{1}{1-\mu}} I,
\end{align*}
\]

where \( \kappa \) is the Lagrange multiplier. After substituting these two functions into the expression for \( I \), we obtain \( p_A \) in expression (20) in the text. Finally, let \( C_A \) and \( C_B \) denote consumption of the final goods \( A \) and \( B \) so that \( \delta \equiv C_A/(C_A + C_B) \) is the consumption share of the final good \( A \) in total consumption. From this follows the CPI given by expression (21) in the text.

A.7 Good B as non-traded

Taking all goods as traded considerably facilitates the model, as demand for all goods is simply the exogenously given world price; thus output and employment levels are determined by the corresponding points in the respective supply curves. The main implication of treating good B as non-traded is that its price would be endogenously determined. To calculate it, we would need to track income levels (from quasi-rents and wages) throughout the economy, introduce a demand system for \( A \) and \( B \), and then solve an additional excess demand equation for B that would yield \( p_B \) jointly with wages and employment levels. It is straightforward to see that in this case changes in employment in the B sector resulting from changes in VA or CSI + NCSI taxes and subsidies as the supply curve of B shifts would be partly offset as \( p_B \) changes in the opposite direction. However, with a downward (upward) sloping demand (supply) curve for B, such price changes would not revert the sign of the employment changes predicted by our model. The result would be that the reductions in employment in the B sector associated with the transit from CSI + NCSI to USI would be smaller. In turn, the expansion of the tax base in the \( (I_1, I_2, A) \) sector would be smaller as well, as would be the additional revenues from the VAT reform. But as discussed in the text, the lower bound on such estimates occurs when in fact the VAT rate increases while CSI + NCSI taxes and subsidies remain constant (since labor re-allocations flow in the direction of higher informality in response to the higher average VAT). These lower bounds have been presented in tables 12 and 13. Given this, there is little gain by the additional modeling required to make \( p_B \) endogenous.
Appendix 2. Data and model calibration

The reference year for all variables is 2008. Data on the government’s fiscal balance is taken from official fiscal accounts (Cuenta de la Hacienda Pública Federal); data on employment from the Economic Census of 2008 together with the National Survey of Occupation and Employment (Encuesta Nacional de Ocupación y Empleo, ENOE) and registries of IMSS. Sources for other data are specified below.

1. Employment

We use two sources of data for employment, depending on the purpose of analysis.

1.1 Table 2: total private occupied labor force, 2008

To present a complete and consistent distribution of the total occupied labor force by size and formality status we use the ENOE, given that the Census only provides information on employment in fixed establishments. The total occupied labor force reported in the ENOE in the second quarter of 2008 is 43,866,696 workers. To identify within this total the portion that is also captured in the Census, we analyzed the characteristics of the workplace that individuals reported in ENOE and compared them with the characteristics of the establishments captured by the Census. For this exercise we identified activities at the 3 digit level using the ENOE and characteristics of the establishment (whether it was fixed or not). After mapping these characteristics we obtained a very similar number of workers in fixed, non-governmental urban establishments: 20,116,834 workers in the Census, and 20,254,726 in the ENOE.82

In addition to the 20,116,834 workers, the Census also identified 4,836,346 workers in government and religious activities. We found a very similar number in the ENOE: 4,645,104. Next, in the ENOE we identified 5,638,429 workers in rural areas (not captured in the Census). As a result, the remaining 13,456,430 workers captured by the ENOE worked in locations that are not fixed (selling in the streets, working as domestic servants, etc.) and were therefore excluded from the occupied labor force captured in the Census. Finally, we divided workers between informal and formal depending on whether they reported access to social security through their jobs, and classified them by firm size as captured in the ENOE.

1.2 Table 8: composition of employment for model calibration

For model calibration we use data from the Census and IMSS registries and not from ENOE because, first, the ENOE is a household survey; for obvious reasons the information from a census is preferred. Second, the size distribution of firms reported in ENOE is not as precise as the one provided by the Census; as discussed below, this information is crucial to estimate the distribution of capital among firms. Third, ENOE classifies economic activities at the three-digit level; in contrast, the Census reports information at the five-digit level. This allows for a more accurate classification of workers between intermediate sectors 1 and 2. In these calculations we exclude government employees and people engaged in religious activities for two reasons. First, the model considers only profit maximizing firms. Second, the proposal to

---

82 Table 2 only reflects 19,629 thousand workers because in this table the occupied population is classified by firm size, and the firm size variable in the ENOE was only defined for that population. The remaining weighted observations have missing values.
reform social insurance is designed for workers in the private sector only (those covered by Chapter A in Article 123 of Mexico’s Constitution). As indicated, 43.86 million workers were occupied in the second quarter of 2008. Subtracting public sector employees and people engaged in religious activities from this number—which according to the Census are 4.83 million workers—leaves 39.03 million workers that need to be distributed among the intermediate and final good B sectors.

The distribution of workers between intermediate sectors 1 and 2 is taken from the Census, which captures 20.12 million workers, once government employees and people engaged in religious activities are excluded. Out of this number, 5.28 million work in sector 1 and 14.84 million in sector 2. In terms of the model, the remaining 18.91 million workers (=39.03 - 20.12) are allocated to the final good B sector.

The 20.12 million workers included in the Census are considered salaried and classified as either formal or informal. Registries from IMSS report 14.18 million workers affiliated during 2008. However, as the Census does not include workers in agricultural, hunting, livestock and forestry activities among others, employees in such sectors must be excluded from the IMSS registries. This leaves a total of 12.76 million workers registered at IMSS, which is then the total number of formal workers in the model. The remaining workers (7.36 million) are classified as informal.

2. Technology

For the case of intermediate and final good B sectors, \( a \) is set to 0.65, consistent with the results of García-Verdú (2005). For final good sector A, \( a_m \) is set to 0.48, which is the average share of intermediate goods in gross output once the production of the household informal sector is taken into account, according to data from National Statistics Office (INEGI) for the period 2003 – 2007. Parameters \( A_1, A_2, \) and \( A_B \) are set at 1874.2, 1837.5 and 219.5, respectively, so that the number of workers replicate the total of formal, informal, and self-employed and family workers in the data. Given these values, parameter \( A_m \) is chosen to scale up activity to replicate the GDP observed in 2008, and is set at 158.1. On the other hand, the national accounts indicate that sector 1 represents approximately 30 percent of total consumption of intermediate goods, which corresponds to parameter \( \gamma \) in equation (A1.10). To the best of our knowledge, there are no estimates available in the literature for the elasticity of substitution between intermediate sectors 1 and 2. Presumably, this elasticity is relatively low and we arbitrarily set \( \mu = -4 \), so that the corresponding elasticity is – 0.20.

3. Price indices

Prices \( p_1, p_2 \) and \( p_B \) are exogenous given our small open economy assumption; we set each to 1/3 to comply with our normalization condition. Parameter \( \delta \) for the CPI in equation (21) in the main text is fixed to 0.9. This implies that relative consumption of goods from the final sector B, \( 1 - \delta \), is consistent with the share of the household informal sector in total

---

83 Intermediate sector 1 includes workers related to: animal breeding, fishing, food and pharmaceutical industries; wholesale and retail trade of food, books, magazines and agricultural goods; passenger transportation; postal services; book and magazine publishing; educational services; and health care and social assistance services.
output, according to national accounts and information from INEGI’s Cuentas Satélite del Subsector Informal de los Hogares.

4. Capital stock

Next we describe how the distribution function for capital $K_z$ in each sector is calculated. This distribution is critical to derive the allocation of labor across firms of different sizes. For these purposes, a method similar to Guner et al. (2008) and Leal (2010) is followed. In particular, capital is assumed to follow a truncated Pareto distribution of the form:

\[
F(K_z) = \frac{1 - \left(\frac{K_{z,\text{min}}}{K_z}\right)^{s_z}}{1 - \left(\frac{K_{z,\text{min}}}{K_{z,\text{max}}}\right)^{s_z}},
\]

where $s_z > 0$ is a shape parameter associated to the distribution in sector $z$, with $K_{z,\text{min}} \equiv K_z$ and $K_{z,\text{max}} \equiv K_z$. The shape parameter is allowed to differ between sectors in order for the model to better fit the data. It turns out that the truncated Pareto distribution is able to explain most of the employment in each intermediate sector with a total mass $1 - f_{z,\text{max}}$. The remaining employment share (which corresponds to the right tail of the distribution with mass $f_{z,\text{max}}$) is obtained by selecting an arbitrary value $K_{z,\text{top}} > K_z$.\footnote{Under the benchmark, the corresponding mass $f_{z,\text{max}}$ in sectors 1 and 2 is 7.8e-06 and 4.8e-05, respectively.} Hence, the distribution of capital has two parts: the bottom side, which accounts for most of the employment, is defined by a truncated Pareto distribution. In contrast, the top side is captured by an extreme value of physical capital. This approach helps to have a better fit for the share of employment in the upper tail of the distribution. The results of this calibration are shown in Panel C of table 8 in the main text. The shape parameter values are fixed to $s_1 = 0.9$ and $s_2 = 0.75$ across all simulations. In general, the model does a good job in replicating the employment shares found in the data, including the values at the tail of each distribution.

5. Taxation, spending and evasion

Consider now the parameters related to taxation, evasion and government revenues and expenditures. In the model intermediate goods sector 1 represents the food and medicines sector of the economy; thus $\tau_1^{VAT} = 0$. On the other hand, intermediate sector 2 represents economic activity subject to the general VAT rate, which in 2008 was 15 percent; thus $\tau_2^{VAT} = 0.15$.\footnote{The statutory tax for these goods in the border Mexican states was 10 percent in 2008. The model abstracts from this geographical dimension and simply sets $\tau_1^{VAT}$ to 15 percent. Starting 2010, the general VAT rate was raised to 16 and 11 percent for non-border and border states, respectively.} The corporate income tax rate $\tau^{IT}$ works like a lump-sum tax in the model. Its value is calibrated so that government revenue out of this tax replicates the data; this implies $\tau^{IT} = 0.106$.\footnote{As a reference, the statutory income tax rate was 0.28 in 2008. In 2010 the rate was increased to 30 percent.} With regards to penalties, the Social Security Law implies that the penalty imposed to a firm caught evading social security contributions is 150 percent of unpaid contributions; thus $\sigma^{CSI} = 1.5$. For the case of VAT and income taxes, the penalty imposed varies between 150 and 170 percent according to the Federal Fiscal Code, but may be higher if they are paid with delay or if there is a previous record of non-compliance with the law; on the other hand, the
penalty may decrease if it is paid promptly. Given the complexity of such scheme, penalties are simply set to 150 percent of the amount evaded, implying $\sigma^{VAT} = \sigma^{IT} = \sigma = 1.5$. Parameter $\lambda$ in the probability of detection function $\lambda(K_z)$ is fixed to 2.14 so that VAT revenues match the data. Differently put, given the tax base and the tax rate, we endogenously calculate the level of evasion that is consistent with the revenues observed in 2008. Similarly, the values for $\nu = 0.1$ and $\lambda^{CSI} = 0.012$ in function $\lambda^z_{CSI}(L_{it}, K_z)$ were chosen so that government revenue out of CSI taxes replicates the data, given the level of evasion of these taxes. The exogenous components of revenue and expenditures from equation (22) in the main text are calculated as the difference from total revenues and expenditures in 2008 obtained from the official fiscal accounts, and the revenues from CSI, VAT and IT taxes, on one hand; and expenditures in subsidies for CSI and NCSI, on the other. This implies that $\overline{R} = 1,852$ and $\overline{G} = 2,679$ thousand million pesos.

6. Social insurance

Parameters related to the valuation of CSI and NCSI, $\beta^{CSI}$ and $\beta^{NCSI}$, are taken from the econometric estimates of Levy (2008), equal to 0.3 and 0.85, respectively. On the other hand, table A2.1 shows that the average tax rate on formal labor, $\tau^{CSI}$, is 38 percent, with government subsidies representing 6 percent (allocated by Law to health and life and disability insurance and retirement pensions); accordingly, $\theta = 0.16$. These values imply that the average per worker subsidy for CSI, $\theta^{CSI}$, was 5,062 pesos. Note that state taxes on salaried labor are included as part of CSI taxes even though the revenues are not allocated to SI. This is because these payments are bundled with other CSI taxes, are exclusive to salaried work, and must be absorbed by firms and workers. The corresponding rate is set at 2%, which corresponds to its statutory value. On the other hand, per worker subsidies for NCSI are calculated by dividing total government subsidies for NCSI programs in 2008, as identified in table A2.2, by the estimated number of informal workers. We list NCSI programs for 151,505 million pesos, which divided by 26.8 million informal workers implies that $\tau^{NCSI} = 5,652$ pesos per worker. Note that public subsidies for health services include spending by sub-national governments, as according to the General Health Law they must contribute along with the Federal Government to financing health services for workers without CSI coverage, and are part of the benefits received by informal workers. State spending in health comes from revenue sharing formulas which are contained in the Federal Budget (under participaciones or aportaciones). In principle we should also have added sub-national spending for pension, housing and day care programs (which are significant), but we were unable to put together a data set for all states and municipal governments. We list NCSI programs by their Spanish name to link directly to the Federal Budget.
### Table A.1 CSI taxes and subsidies*

<table>
<thead>
<tr>
<th></th>
<th>Total rate</th>
<th>Government subsidy</th>
<th>Firms and worker contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health insurance</td>
<td>17.5</td>
<td>4.50</td>
<td>13.00</td>
</tr>
<tr>
<td>Life and disability insurance</td>
<td>2.75</td>
<td>0.25</td>
<td>2.50</td>
</tr>
<tr>
<td>Work-risk insurance</td>
<td>1.75</td>
<td>0</td>
<td>1.75</td>
</tr>
<tr>
<td>Retirement pensions</td>
<td>8.00</td>
<td>1.70</td>
<td>6.30</td>
</tr>
<tr>
<td>Day care services</td>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Housing</td>
<td>5.00</td>
<td>0</td>
<td>5.00</td>
</tr>
<tr>
<td>State labor taxes</td>
<td>2.00</td>
<td>0</td>
<td>2.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>( \tau^{CSI} = 0.38 )</td>
<td>( \theta \tau^{CSI} = 0.06 )</td>
<td>( (1 - \theta) \tau^{CSI} = 0.32 )</td>
</tr>
</tbody>
</table>

*Calculated as a % of the average formal wage as some contribution rates vary with wage levels; see IMSS (2005).

### Table A.2 Subsidies to NCSI programs*

<table>
<thead>
<tr>
<th><strong>Health</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fassa (Ramo 33)</td>
<td>48,480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Spending for health services</td>
<td>24,715</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seguro de Salud para la Familia (IMSS)**</td>
<td>1,115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seguro Popular</td>
<td>36,250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMSS-Oportunidades**</td>
<td>6,370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitales de Especialidad (Ramo 12)</td>
<td>12,416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seguro Universal de Primera Generacion</td>
<td>1,699</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total health</strong></td>
<td>131,045</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day Care</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estancias Infantiles (Sedesol)</td>
<td>1,711</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Housing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fonhapo</td>
<td>2,342</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conavi</td>
<td>4,984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitat</td>
<td>1,887</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total housing</strong></td>
<td>9,213</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pensions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adultos Mayores de Setenta</td>
<td>9,536</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>151,505</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*million pesos. **These two programs, although operated by IMSS, provide health services for informal workers financed directly by the Federal Government. Despite the somewhat confusing name, the program IMSS-Oportunidades is different from the targeted poverty program Oportunidades, which is excluded from these accounts.
Appendix 3: Costs of USI

Per worker annual costs of USI of 14,330 pesos consist of three components: 10,118 for health insurance, 3,492 for retirement pensions, and 720 for life and disability pensions.

1. Costs of health insurance

As discussed in the text, we take as reference the IMSS health package. In 2008 IMSS received 109.3 thousand million pesos (tmp) in CSI taxes from firms and workers, and 44.3 tmp in CSI subsidies from the government, for a total expenditure of 153.7 tmp; see table 7 in the text. As described in IMSS (2005), however, IMSS uses part of its revenues to pay for the pension liabilities of its own already retired workers (the Regimen de Jubilaciones y Pensiones, or RJP), not to provide health services to affiliated workers. In 2008 16 cents of every peso of IMSS revenues were diverted to the RJP, implying that actual expenditures in health services per worker were (153.7)(0.84)/12.76 = 10,118 pesos, given an estimate of 12.76 million formal workers in that year. Given a total occupied labor force of 39.03 million workers, the total costs of health insurance under USI are 394.9 tmp. On the other hand, the RJP pension liabilities would have to be paid directly by the Federal Government, for a total amount of 27 tmp in 2008 (including the payments to RJP from the other insurance items administered by IMSS), which is the amount added to the costs of the proposal in table 12.

Table A3.1 compares health expenditure under CSI + NCSI vs. USI.

Table A3.1: Annual subsidies per worker for health insurance

<table>
<thead>
<tr>
<th></th>
<th>CSI*</th>
<th>NCSI</th>
<th>USI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm and worker contributions</td>
<td>7,166**</td>
<td>0</td>
<td>10,118</td>
</tr>
<tr>
<td>Government subsidies</td>
<td>2,952</td>
<td>4,904</td>
<td>10,118</td>
</tr>
<tr>
<td>Total</td>
<td>10,118</td>
<td>4,904</td>
<td>10,118</td>
</tr>
</tbody>
</table>

*After deducting contribution for RJP; **the incidence of which falls 70% on workers and 30% on firms.

Note that under USI spending in health for informal workers increase by 106% while, by construction, stays constant for formal workers. As a result, resources for health insurance for workers increase from 260.1 tmp (129.1 tmp from CSI net of RJP, and 131 tmp from NCSI) to 394.9 tmp under USI, a 52% increase, which in turn implies channeling an additional 1.1% of GDP to public health services under our proposal.

2. Retirement and life and disability pensions

In 2008 two minimum wages was equivalent to a monthly salary of 3,000 pesos. Contributions by firms and workers to retirement pensions for a worker in that salary were 199 pesos. Adding 93 pesos from the contribution made by the government results in total contributions of 292 pesos a month, or 3,492 pesos a year. Multiplying this by 39.03 million workers yields a total cost of 136.3 tmp.

IMSS (2005) shows that the actuarial cost of life and disability insurance net of IMSS RJP costs (which, as described above, are absorbed by the Federal Government) is 2% of the wage. For a worker earning two minimum wages the costs of life and disability insurance is then 60 pesos a month, or 720 pesos a year. Multiplying this by 39.03 million workers yields a total cost of 28.1 tmp. Adding this to the 394.9 tmp for health and 136.3 tmp for retirement pensions yields the total cost of USI of 559.3 tmp included in table 12 in the text.
References


90


