Informality, Inflation and Fiscal Progressivity in Developing Countries

October 13, 2025

Daniel Jaar[1], João Ritto[2]

Abstract

We develop a dynamic general equilibrium model with heterogeneous households and a cash-intensive informal sector that replicates two empirical patterns: the negative relationship between informality and firm productivity, and the declining share of informal consumption with household wealth. The non-homotheticity of informal consumption implies that tax incidence is heterogenous: poor households pay less consumption taxes but are more exposed to inflation. We use the model to study the distributional effects of financing government revenue through seigniorage versus consumption taxes. Calibrated to Peru – where informality accounts for around half of economic activity – the model shows that informal purchases provide significant savings through lower prices, particularly for poor households, who save up to 11% compared to purchasing the same bundle formally. The model also uncovers substantial variation in preferences over revenue-neutral combinations of inflation and consumption taxes: households in the top expenditure decile would like inflation to be as high as 12%, while those in the bottom favor inflation below 5%. This disagreement grows with the size of the informal sector.

1 Introduction

Governments in developing countries face challenges in raising revenue, collecting only about half as much as rich countries as a share of GDP (Besley and Persson, 2014). The presence of large informal sectors that evade taxation is understood as one of the major factors limiting

[1]Department of Economics, European University Institute.

[2]Department of Economics, University of Toronto, corresponding author, email: j.ritto@utoronto.ca

We are grateful to Harold Cole, Joachim Hubmer, Dirk Krueger, Benjamin Lester, José-Víctor Ríos-Rull and Guillermo Ordoñez for invaluable guidance. We also thank Wai-Ming Ho, Sergio Ocampo Díaz, Stephen Snudden and seminar participants at the University of Pennsylvania, Annual Workshop of Southern Ontario Macro Economists (2025), and the 59th Annual Meeting of the Canadian Economic Association for helpful comments.

the effectiveness of conventional tax instruments such as income and consumption taxes, and has been pointed as one of the reasons why developing countries sustain higher levels of long-run inflation Aisen and Veiga (2008). Traditional taxation incentivizes informality by making formal participation more costly. Inflation, by contrast, functions as a tax on cash holdings – capturing revenue from both formal and informal transactions (Nicolini, 1998). Quantitative analyses of optimal taxation in the presence of informality can rationalize the empirical positive relationship between inflation and informality found across countries (Koreshkova, 2006; Aruoba, 2021).

However, the nature of the informal economy implies that policies disproportionately affecting a sector over the other may have significant distributional consequences. On the employment side, informal firms are typically owned by and employ low-skilled workers (La Porta and Shleifer, 2014; Ulyssea, 2018). On the consumption side – our focus in this paper – poorer households purchase a larger proportion of their consumption bundle from informal firms. Indeed, in a recent paper, Bachas et al. (2024a) document that the share of expenditure in informal establishments steeply declines with income in a broad set of countries, which they term as a (negatively sloped) informality Engel curve. Because informal firms avoid taxation, these consumption patterns imply that richer households bear the brunt of consumption taxes. In turn, we argue that the cash-intensive nature of informality leads poor households to hold a larger share of their wealth in cash, making them more vulnerable to inflation.

In this paper, we quantitatively assess the distributional effects of public finance choices in economies with large informal sectors, with a focus on seigniorage and consumption taxes. To achieve this, we develop a dynamic general equilibrium model with heterogeneous households and firms. The key innovation is that the model incorporates both a formalization decision by firms, involving formalization costs and the obligation to pay consumption taxes, and household decisions over where to shop (formal vs informal) and how to pay (cash vs non-cash). Our modeling of the latter is built on a rich description of the goods market – with households buying different varieties of different quality – that fits tractably into a dynamic model and implies that the informal share of a household's consumption bundle, as well as her money demand, are endogenous and affected by policy choices. Moreover, this formulation enables the model to reproduce two salient empirical patterns: the positive relationship between firm size and formality, and the negative relationship between household wealth and informal consumption.

In our framework, the informal sector's ability to avoid taxes and regulations results in lower prices which benefit all consumers, but especially poorer households, who disproportionately purchase lower-quality goods from informal firms. Our first quantitative result is that total savings due to informal purchases – what we term the *informal discount* – are substantial and unequally distributed. On average, households save about 7%; those in the lowest expenditure decile save roughly twice as much as those in the highest. Second, because informal activity is cash-intensive, inflation erodes some of the progressivity achieved from taxing formal-sector consumption. The lowest income decile pays an effective consumption tax rate equal to 27% of that of the highest decile, a ratio that rises to 41% once inflation's effect on the overall effective tax rate is considered. Third, the differing tax incidences generate sizable heterogeneity in preferences for revenue neutral combinations of consumption taxes and inflation: the richest decile prefers an inflation rate of 12% and consumption tax rate of 13%, whereas the poorest decile prefers 4% inflation and 18% consumption tax rate.

The model features sorting into informality, by both households with heterogeneous wealth and firms with heterogenous productivity, as the key mechanism that allows us to match the empirical patterns regarding informality. The goods market features a continuum of varieties differentiated by quality. Households consume one unit of each variety, choosing both its quality and whether to purchase it from the formal or informal sector. They have no intrinsic preference for either sector and buy from the cheaper source for a given quality. On the production side, firms differ in productivity and choose which variety to produce, the quality composition of their output, and whether to formalize. Informal firms avoid regulatory costs but must pay tax evasion costs and are restricted to cash transactions to avoid detection, following Aruoba (2021). We assume enforcement is more effective for higher-priced goods, making informal sales of high-quality products more costly. As a result, in equilibrium, low-quality goods are cheaper in the informal sector, while high-quality goods are more affordable in the formal sector. Additionally, the presence of formalization costs implies that only sufficiently productive firms formalize even though formalization yields higher per-unit profits, consistent with the evidence (La Porta and Shleifer, 2014; Ulyssea, 2018).

From the household's perspective, the total purchasing cost of a good depends not only on its sector and quality but also on the method of payment. Paying with cash entails a loss from forgone interest due to a cash-in-advance constraint, while using credit – available only for formal purchases – incurs financial service fees as in Erosa and Ventura (2002). Poorer households, who tend to consume lower-quality goods, purchase a larger share of their bundle from the informal sector and hold more of their wealth in cash – both patterns consistent with the data. This endogenous, non-homothetic demand for informality generates variation in effective tax burdens across the income distribution and leads to heterogeneous preferences over revenue-neutral fiscal reforms.

To evaluate the quantitative implications of our theory, we calibrate the model so that it matches key aggregate and distributional facts about informal economic activity in Peru, a representative developing country with a large informal sector. In particular, the model replicates the size of Peru's informal sector, the negative relationship between expenditure and the share of informal consumption, and recent estimates of the passthrough of consumption taxes to both formal and informal prices (Bachas et al., 2024a). We use the calibrated model to study the aggregate and distributional consequences of revenue-neutral changes in long-run inflation and consumption taxes. Increasing inflation from 4% to 10% allows for a reduction of the consumption tax rate from 18 to 13% which contracts the informal sector and benefits the top 50% of the wealth distribution at the expense of the bottom 50%. Importantly, our distributional results are not driven by costly credit provision but rather the non-homothetic demand for informal consumption. Indeed, eliminating costly credit provision does not substantially change the dispersion in informal discounts nor in the preferences for inflation. Finally, repeating our quantitative exercises in economies with varying degrees of informality reveals that the distributional considerations of inflation are higher in countries where informality is more prevalent.

Our paper contributes to the body of work studying how informality limits governments' capacity to raise revenue and shapes the distributional impacts of fiscal policy in emerging economies (Besley and Persson, 2014; Bachas et al., 2024b). Existing research has examined this issue primarily through the lens of social security contributions to finance social security or unemployment insurance.[3] Ulyssea (2018) finds that stricter enforcement of informal firms lowers the wage of low-skilled workers in Brazil, while Asatryan and Gomtsyan (2020) shows that enforcing VAT payment among large retailers disproportionately hurts poorer consumers in Armenia. To the best of our knowledge, ours is the first paper to quantify how informal firms' ability to bypass regulations impacts household's purchasing power through lower prices. Our focus on seigniorage brings us closer to the literature on the heterogeneous welfare costs of anticipated inflation which studies different mechanisms that induce heterogeneous money holdings across the income distribution in the context of developed countries (Erosa and Ventura, 2002; Cao et al., 2021; Cirelli, 2022). We show that the differential demand for informal consumption is sufficient to generate large differences in money holdings across the expenditure distribution, with significant distributional considerations.

We borrow extensively from Bachas et al. (2024a), who study the equity implications of consumption taxation (and consumption tax exemptions) in developing countries. They show that the non-homothetic informal consumption patterns are consistent with richer households valuing the higher quality goods sold by formal retailers, rather than differences in access to formal firms or other observable characteristics. Using a structural framework, they show that the optimal tax differentiation between food and non-food goods is increasing with development, and not justified for low income countries on equity grounds. In line with their findings, our model creates differences in the formality composition of consumption through differences in the prices of goods of different quality levels. We focus on the tradeoff between inflation and consumption taxes in a dynamic general equilibrium model and show that the progressivity achieved through consumption taxes is partially undone by the regressive consequences of inflation.

Our work is also related to the literature studying the public-finance motive for inflation in economies with large shadow economies using the Ramsey approach (Nicolini, 1998; Koreshkova, 2006; Aruoba, 2021). These papers find that the need to tax the informal sector implies that the Friedman rule is no longer optimal, and that it can rationalize the positive relationship between inflation and informality observed in the data. We depart from the representative agent abstraction to study the impact of public finance decisions across the wealth distribution in light of the facts about heterogeneity recently documented in the empirical literature. We assume that informal firms must transact in cash to avoid government oversight.

[3] Examples are Cirelli et al. (2021); Jaar (2024); Joubert (2015), among others.

[4] The implications of the exclusion of the informal sector from the financial system and its connection with tax evasion have been studied both theoretically and empirically in Aruoba (2021); Caballé and Panadés (2004); D'Erasmo and Moscoso Boedo (2012); Narayanan et al. (2020).

The paper proceeds as follows. Section 2 describes and characterizes the model, Section 3 presents our calibration strategy, Section 4 presents our results and Section 5 concludes.

2 Model

Many of the model ingredients we use are common to Koreshkova (2006). We introduce heterogeneity on the household side by incorporating uninsurable productivity shocks and a structure that delivers the essential non-homotheticities in the consumption bundle. As in Aruoba (2021) we assume that informal purchases must be done using money.

2.1 Setup

Environment

Time is infinite and discrete. The agents in the economy are the households, firms, and government. Households choose how much to work and save, as well as the quality and formality composition of their consumption bundle. Firms employ households and must choose which variety and in which sector to produce, as well as the quality composition of their production. The government supplies money and government bonds and must finance an exogenous expenditure stream through consumption taxes on the formal sector and seigniorage.

There is a continuum of goods varieties indexed by $j \in [0, 1]$. Each variety can be produced by formal and informal firms. Producers of variety j can produce goods of different quality $q > 0$ using labor as their only input. Varieties differ in two dimensions: the ease with which firms can hide their trade from the authorities and the cost consumers must incur in order to purchase this good with credit rather than cash. In particular, the government's monitoring ability is increasing in j, and the cost of using credit is decreasing in j. This means that varieties with a low j index are easier to hide from the government, but more costly to trade using credit, while the opposite holds for varieties with a high j index.[5]

The real interest rate is exogenous and given by $r_t = r^*$. This can be rationalized by a supply of risk-free assets composed not only of government bonds but also of external debt issued elastically by foreigners at a fixed real interest rate.

[4]We don't require that informal consumption be financed exclusively with M0, but rather that it cannot be financed with financial instruments that allow individuals to hedge against inflation, such as credit cards.

[5]Implicitly, we assume that the underlying informational friction that makes the transaction/production of certain varieties difficult to monitor for the government also makes the provision of financial services costlier.

Households

The economy is populated by a unit measure of households. Each household starts period t with real wealth x_t, labor productivity ϵ_t and must choose how much to work, how much of its wealth to hold in cash and the interest-bearing risk-free asset, as well as three characteristics of its consumption bundle: the quality of the good it purchases of each variety, whether to buy it from the formal (f) or the informal (n) sector, and whether it finances each purchase using cash or credit.

Households maximize expected discounted lifetime utility:

$$\mathbb{E}_t \sum_{k=0}^{\infty} \beta^k u(q_{t+k}, \ell_{t+k}).$$

Here, β is the discount factor, ℓ is the labor supply, and period utility function u is strictly increasing (decreasing) in its first (second) argument and concave. Composite q is a consumption aggregator over varieties given by

$$q = \inf_j \Big\{ \max\{q_{jf}, q_{jn}\} \Big\},$$

where q_{jf} and q_{jn} denote the quality of formal and informal purchases of variety j. The maximum operator represents preferences in which formal and informal goods of a variety j are perfect substitutes and the household derives no utility from consuming two goods of the same variety. The specific aggregator chosen implies different varieties are perfect complements and the household will choose to purchase the same quality for every variety.

In period t, a household that ended the previous period with real wealth x_t chooses how much cash $m_t \geq 0$ and interest-bearing assets $b_t \geq 0$ (both in real terms) to hold, subject to the beginning of period budget constraint:

$$m_t + \frac{b_t}{1+i_t} = \frac{x_t}{1+\pi_t},$$

where π_t is the inflation rate that depreciates the value of the household's wealth from one period to the next, and i_t is the nominal interest rate paid on the interest-bearing asset. In its portfolio decision, the household considers that it must carry enough cash to pay for its informal consumption and formal consumption purchased with cash. This is described by the following cash-in-advance constraint:

$$\int p_{jnt}(q_{jnt})\, dj + \int \mathbb{I}\{j \in \mathcal{J}^m\}\, p_{jft}(q_{jft})\, dj \leq m_t, \tag{CIA}$$

where $p_{jnt}(q)$ and $p_{jft}(q)$ are the prices of a good of variety j with quality q purchased in the informal and formal sectors, and $\mathcal{J}^m$ is the subset of varieties purchased formally that are bought using money.[6] As long as the nominal interest rate is positive the return of bonds

[6] Note that if $q_{jft} > 0$, then $q_{jnt} = 0$ and $p_{jnt}(q_{jnt}) = 0$.

dominates that of money, so constraint CIA will bind. In turn, varieties purchased formally using credit require the payment of the financial service cost $\gamma(j)$.

The evolution of a household's assets from period t to period $t+1$ is given by:

$$x_{t+1} = \; m_t + b_t + \epsilon_t \ell_t - \int p_{jnt}(q_{jnt})\, dj - \int p_{jft}(q_{jft})\, dj - \int \mathbb{I}\{j \in \mathcal{J}^c\}\, \gamma(j)\, dj,$$

where $\mathcal{J}^c$ is the subset of varieties purchased formally that are bought using credit. Labor productivity follows a Markov process with an unconditional mean equal to 1. Because we make the wage per efficiency unit of labor the numeraire, labor income (in these units) is $\epsilon_t \ell_t$. We assume that formal good purchases are paid after the household receives payment on its government bond holdings, but before it receives its labor income. This implies the following borrowing constraint:

$$x_{t+1} \geq \epsilon_t \ell_t.$$

That is, buying with credit provides a way to avoid forgoing the nominal interest rate by holding cash, but it does not constitute a way of borrowing against the labor income paid at the end of the period when x_t is low.

Firms

Each period a measure of potential firm entrants decides whether to participate in the economy for a single period, which involves paying an entry cost κ_e. Upon entry, firms draw a productivity $z \sim \Psi$ and must choose a variety j and a sector $s \in \{f, n\}$ in which to produce. Formalization entails the payment of a fixed cost κ_f (common across varieties) which subjects the firm to taxation but brings the possibility of selling goods with credit. Informal firms avoid these costs, but may only accept cash and face tax avoidance costs.

Firms operate according to a decreasing returns to scale production function that uses labor ℓ as the only input: $F(z, \ell) = z\ell^{\alpha}$ $(\alpha < 1)$. Labor can be hired in a competitive labor market. Given the firm's output $F(z, \ell)$ each firm decides how many goods of each quality to produce. This transformation is linear: producing one good of quality q requires q units of output. This implies that the firm may produce $(s(q))_q$ units of each quality subject to the following capacity constraint:

$$\int q s(q)\, dq \leq F(z, \ell). \tag{CC}$$

Conditional on labor input ℓ, the firm's revenue maximization problem consists in choosing how to allocate its capacity $F(z, \ell)$ into the production of goods of different quality. Operating revenue in each sector is given by:

$$R_{jf}(z, \ell) = \max_{s(q)} \int \left(\frac{p_{jf}(q)}{1+\tau} \right) s(q) dq \qquad \text{subject to (CC)},$$

$$R_{jn}(z, \ell) = \max_{s(q)} \int \left[p_{jn}(q) - \chi_j(p_{jn}(q)) \right] s(q) dq \qquad \text{subject to (CC)}.$$

Formal firms must pay a sales tax given by τ, whereas informal firms incur tax avoidance costs embodied in function $\chi_j(p)$. We assume that it is harder to hide higher value transactions from the government, hence evasion costs are increasing in the value of the transaction, $\chi_j'(p) > 0$.[7] Additionally, the government can more easily monitor and detect trade of good varieties with a higher index, such that tax avoidance costs are increasing in j: $\chi_k(p) > \chi_j(p)$ for $k > j$. That is, varieties that are cheaper to buy with credit are also more prone to government monitoring.

Firms will choose the sector that leads to the higher profits, defined as operating revenue net of variable and fixed costs:

$$\begin{aligned}\Pi_{js}(z) &= \max_{\ell} R_{js}(z,\ell) - w\ell,\\ \Pi_j(z) &= \max_{\ell} \left\{\Pi_{jf}(z) - (\kappa_e + \kappa_f), \Pi_{jn}(z) - \kappa_e\right\}.\end{aligned}$$

We use $\ell_{js}(z)$ to denote optimal labor demand in each sector. Due to free entry, expected profits must equal zero:

$$\int \Pi_j(z)\, d\Psi(z) = 0. \tag{FEC}$$

Entry, formalization, and tax avoidance costs are paid to a firm that uses labor to deliver these services, with a linear technology. Because the wage is our numeraire, the total costs equal the amount of labor allocated to these non-productive uses.

Government

The government starts period t with liabilities issued in the previous period in the form of outstanding real bond holdings B_{t-1} and real money supply M_{t-1}. It collects the consumption taxes from the previous period T_{t-1}, issues new bonds and money, and must pay for expenditures G_t. Thus, the law of motion for government liabilities is as follows:

$$\frac{B_t}{1+i_t} + M_t + \frac{T_{t-1}}{1+\pi_t} - G_t = \frac{B_{t-1} + M_{t-1}}{1+\pi_t}.$$

Indexing households using $h \in [0,1]$, and q_{hjf} as the quality of the good of variety j purchased by household h in the formal sector, taxes collected are given by:

$$T_{t-1} = \tau_{t-1} \int\int \frac{p_{jf,t-1}(q_{hjf,t-1})}{1+\tau_{t-1}}\, dj\, dh.$$

Government expenditures are purchases of goods from the formal sector:

$$G_t = \int\int \frac{p_{jft}(q)}{1+\tau_t} g_{jt}(q)\, dq\, dj,$$

where $g_{jt}(q)$ is the number of goods of variety j and quality q that the government purchases.[8]

[7] Meghir et al. (2015) and Ulyssea (2018) use similar enforcement cost functions that are increasing in firm size. As in their work, in our model only sufficiently productive (and therefore large) firms will formalize, thus rationalizing the same empirical pattern.

[8] We make government purchases tax-free, so that taxes only include revenue from household purchases.

Market clearing

The markets in this economy are the goods markets (one market per variety-sector-quality combination), the labor market and money market. We describe market clearing briefly and provide detailed explanations in Section A of the Appendix.

Market clearing in the market for goods of variety j, sector s, and quality q requires that the total household (and government) demand for this particular good equals its supply by firms. The money market clears if the money supplied by the government equals the money demanded by households. Because we assume that there is an elastic foreign supply of risk-free assets at real interest rate r^*, the corresponding condition is given by:

$$\mathbb{E}_t \frac{1+i_t}{1+\pi_{t+1}} = 1 + r^*. \tag{AMC}$$

Given that households may hold external debt, consistency requires that such income be used in imports. We let those imports be in the form of labor. Hence, the labor market clears if the number of efficiency units of labor supplied by households, plus the amount of foreign labor it can hire with interest income on foreign assets held, equals the demand for labor by producer firms in both sectors plus the amount of labor allocated to non-productive uses (entry costs, formalization costs, tax avoidance costs, and financial services costs).

In the next subsection, we introduce a series of results that simplify the model's solution and use these to define a stationary equilibrium formally in recursive formulation.

2.2 Model solution

The environment described in the previous subsection has a continuum of market-clearing conditions. However, its structure allows for a significant simplification of its characterization. In particular, two equilibrium objects – the revenue per unit of output in both the informal and formal sectors – are sufficient to characterize the entire equilibrium price schedules, and the goods market clearing can be described by a single condition for each sector. This holds because firms will be indifferent about how many goods of each quality to supply, making quality composition determined in equilibrium solely by the demand side.[9] In turn, using equilibrium price schedules, we characterize the static optimal household shopping behavior separately from the intertemporal decision. This allows us to write a simple recursive representation of the household's problem.

Price schedules and firm behavior

Proposition 1 shows that revenue per unit of output in each sector, (μ_f, μ_n), are sufficient statistics to characterize the equilibrium price schedules. We describe the intuition. Formal proofs are provided in Section B of the Appendix.

[9]Additionally, this means that specifying the quality composition of government purchases is unnecessary.

Proposition 1 *There exist (μ_f, μ_n), representing revenue per unit of output in each sector, such that equilibrium price schedules $(p_{jf}(q), p_{jn}(q))_{j\in[0,1]}$ satisfy:*

1. $p_{jf}(q) = \mu_f(1+\tau)q \ \forall j$,
2. $p_{jn}(q) = \mu_n q + \chi_j\left(p_{jn}(q)\right) \ \forall j$.

These prices make firms indifferent about the quality composition of their production.

To understand the intuition behind Proposition 1, it is useful to focus on a firm producing variety j in sector f. Due to the linearity of the quality transformation, a firm can produce q_1 units of quality q_2 or q_2 units of quality q_1 with the same inputs. For the firm to be willing to sell the two different quality levels, she must be indifferent between the two. That is, the revenue per unit of output from selling q_1, given by $\frac{p_{jf}(q_1)}{(1+\tau)q_1}$, must equal $\frac{p_{jf}(q_2)}{(1+\tau)q_2}$, the equivalent from selling q_2. An analogous logic holds for informal producers and thus, the following equalities regarding variety j price schedules must hold:

$$\frac{p_{jf}(q)}{(1+\tau)q} = \mu_{jf}, \quad \frac{p_{jn}(q) - \chi_j(p_n(q))}{q} = \mu_{jn}, \quad \forall q. \tag{1}$$

Hence, the structure of the revenue maximization problem implies that revenue per unit of output in each sector is independent of quality and given by μ_{js}, which is an equilibrium object that is taken as given by the firm. Free entry into varieties implies that profits equalize among producers with the same productivity z, which requires that μ_{jf} and μ_{jn} are actually independent of j. Figure 1 illustrates the resulting price schedules.

Proposition 2 shows that equilibrium price schedules imply a simple threshold rule for formalization.

Proposition 2 *There exists a productivity threshold $\hat{z}^n$, common across varieties, such that firms remain informal if $z \leq \hat{z}^n$ and formalize otherwise.*

Proposition 1 implies that revenues in sector s are given by $R_s(z,l) = \mu_s F(z,l)$, and we arrive at a simple expression for operating profits:

$$\Pi_s(z) = \max_{\ell} \mu_s F(z,\ell) - w\ell.$$

For both formal and informal firms to coexist in equilibrium, it must be that $\mu_f > \mu_n$: formal firms gain access to a higher μ_s in exchange for paying a fixed cost κ_f. Only productive enough firms will do so, and this threshold is determined by the value $\hat{z}^n$ that solves $\Pi_f(\hat{z}^n) - \Pi_n(\hat{z}^n) = \kappa_f$. That is, in the model informal firms will be less productive than their formal counterparts, in line with the data.

Household consumption behavior

Incorporating results from Propositions 1 and 2 allows us to characterize the household's optimal consumption behavior. Indeed, consider the problem of a household that wants to buy

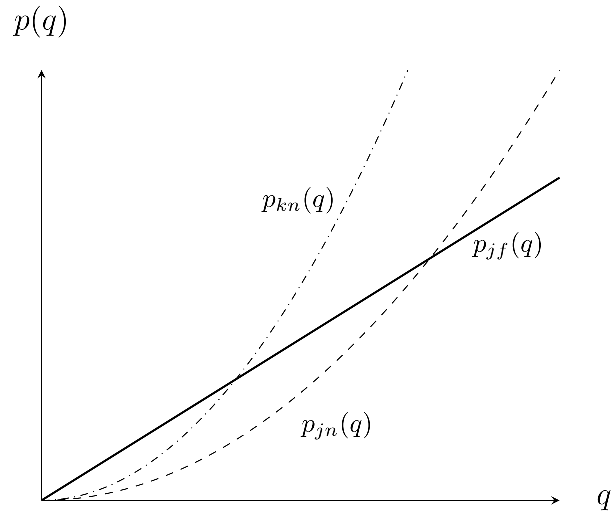


Figure 1
Formal and Informal Price Schedules, $k > j$

a good of variety j of quality q. Given that formal and informal goods are perfect substitutes, optimal sector and method of payment will simply minimize expenditure. Optimal behavior is described in Proposition 3.

Proposition 3 *For each variety j there exist quality thresholds $\hat{q}_j^n$ and $\hat{q}_j^m \geq \hat{q}_j^n$ such that:*

1. *The household buys informally if $q_j < \hat{q}_j^n$ and formally if $q_j \geq \hat{q}_j^n$*
2. *The household pays with cash if $q_j < \hat{q}_j^m$ and uses credit if $q_j \geq \hat{q}_j^m$.*

Thresholds $\hat{q}_j^n$ and $\hat{q}_j^m$ are decreasing in j.

Purchasing the good informally has a direct cost of $p_{jn}(q)$ plus the forgone interest from carrying cash, making the total expenditure from purchasing q informally equal to $p_{jn}(q)(1+i)$. In turn, buying it from the formal sector costs $\mu_f(1+\tau)q$ plus forgone interest of $i\mu_f(1+\tau)q$ if using cash, or $\gamma(j)$ if using credit. Figure 2 shows the expenditure associated with each of the three possible alternatives. It is cheaper to purchase low quality goods from informal firms because evasion costs are low. Because the former increase at an increasing rate, for sufficiently high quality it becomes cheaper to purchase the good formally with money. Finally, the household will buy formally with credit once forgone interest exceeds the fixed cost of purchasing with credit. In other words, Proposition 3 implies that the within-variety expenditure function is given by:

$$e_j(q) = \begin{cases} p_{jn}(q)(1+i) & \text{if } q < \hat{q}_j^n, \\ \mu_f(1+\tau)q(1+i) & \text{if } q \geq \hat{q}_j^n \ \& \ q < \hat{q}_j^m \\ \mu_f(1+\tau)q + \gamma(j) & \text{if } q \geq \hat{q}_j^n \ \& \ q \geq \hat{q}_j^m. \end{cases}$$

Our monotonicity assumptions on $\chi(j)$ and $\gamma(j)$ imply that once it becomes cheaper to buy quality q of variety j formally with cash (or formally with credit), this will be the case for

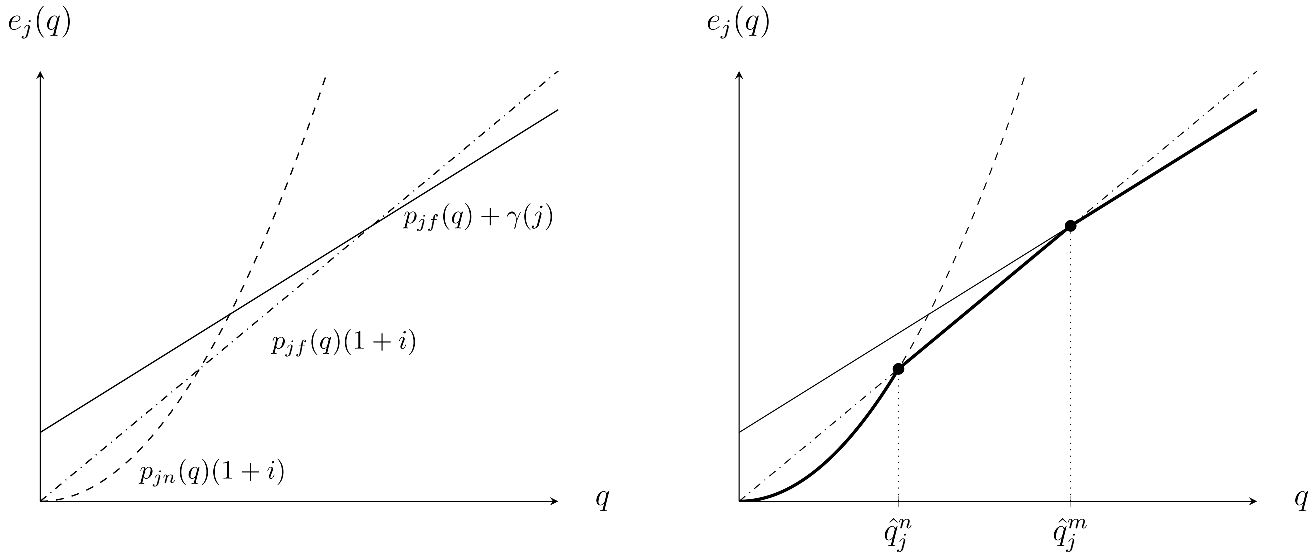


Figure 2
Expenditure function for variety j

all varieties $k > j$. We show that the *quality* threshold rules for sector and method of payment decisions within varieties imply analogous *variety* threshold rules for composite good q.

Proposition 4 *For each quality level $q > 0$ there exist variety thresholds $\hat{j}^n(q),\ \hat{j}^m(q) \in [0,1]$, with $\hat{j}^n(q) \leq \hat{j}^m(q)$, such that:*

1. *The household buys informally the varieties $j < \hat{j}^n(q)$ and formally the remaining.*
2. *The household pays with cash the varieties $j < \hat{j}^m(q)$ and uses credit for the remaining.*

Thresholds $\hat{j}^n(q)$ and $\hat{j}^m(q)$ are decreasing in q.

Proposition 4 means that the model is able to reproduce a key aspect of the evidence documented by Bachas et al. (2024a) – wealthier households, that want to consume higher levels of q, will purchase a larger proportion of their consumption bundle from the formal sector. Additionally, these households will also buy more varieties with credit. Figure 3 illustrates the variety thresholds.

Using Proposition 4 we can write the household's expenditure function and money demand as a function of q:

$$E(q) = \int_0^{\hat{j}^n(q)} (1+i)p_{jn}(q)\,dj + \int_{\hat{j}^n(q)}^{\hat{j}^m(q)} \mu_f(1+\tau)(1+i)q\,dj + \int_{\hat{j}^m(q)}^{1} [\mu_f(1+\tau)q + \gamma(j)]\,dj, \tag{2}$$

$$m(q) = \int_0^{\hat{j}^n(q)} p_{jn}(q)\,dj + \int_{\hat{j}^n(q)}^{\hat{j}^m(q)} \mu_f(1+\tau)q\,dj. \tag{3}$$

Finally, we are able to write the household's problem in recursive form as follows:

$$V(x, \epsilon) = \max_{q \in [0, \bar{q}], \ell \geq 0} u(q, \ell) + \beta \mathbb{E} V(x', \epsilon'), \tag{4}$$

$$st : x' = \left(\frac{1+i}{1+\pi} \right) x + \epsilon \ell - E(q), \tag{5}$$

where maximum consumption level $\bar{q}$ is a consequence of the borrowing constraint $x' \geq \epsilon \ell$ and is given by:

$$E(\bar{q}) = \left(\frac{1+i}{1+\pi} \right) x$$

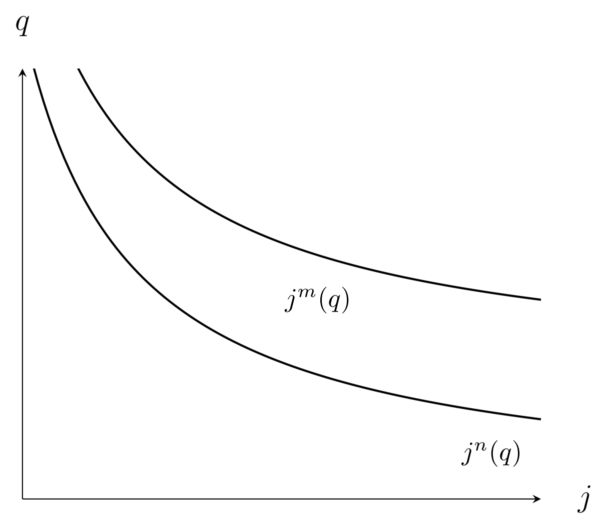


Figure 3
Variety thresholds

Goods market clearing

We have mentioned how the continuum of market-clearing conditions in the goods market can be reduced to a single condition for each sector. Intuitively, because Proposition 1 determines equilibrium price schedules such that firms are indifferent about the quality composition of their supply, the proportion of supply of a certain variety and quality can be made equal to these same proportions for demand. Hence, the only two conditions that need still be imposed are that the aggregate production capacity in each of the two sectors must satisfy the total demand of goods of different quality and variety in that sector. Below we define a stationary equilibrium where these two conditions are outlined. Section B of the Appendix provides a derivation from the original market-clearing conditions.

Stationary equilibrium

A stationary equilibrium in this environment consists of **i)** revenues per unit of output in the two sectors, (μ_f, μ_n), and inflation rate π; **ii)** household intertemporal policy functions $q(x, \epsilon)$

and $\ell(x,\epsilon)$, and intratemporal decisions $\hat{\jmath}^n(q)$, $\hat{\jmath}^m(q)$ and $m(q)$; **iii)** firm policy functions for labor demand $(\ell_f(z), \ell_n(z))$ and formalization threshold $\hat{z}^n$; **iv)** government policy mix (B, M, i, τ, G, g_N), where B (M) is the amount of real bonds (real money) and g_N is the nominal growth rate of both of these liabilities; **v)** a household distribution across individual states $\Phi(x,\epsilon)$ with its law of motion $\mathcal{T}$ and a mass of entrants ν^F such that:

1. Given price schedules implied by (μ_f, μ_n) as described in Proposition 1, policy functions $q(x,\epsilon)$ and $\ell(x,\epsilon)$ solve 4. Moreover, sector-of-purchase and payment-method decision thresholds $\left(\hat{\jmath}^n(q), \hat{\jmath}^m(q)\right)$ are consistent with Propositions 3 - 4 and money demand $m(q)$ follows condition 3.
2. Firms maximize profits and the free entry condition FEC holds.
3. (B, M, τ, G, g_N) satisfy the law of motion for government liabilities and inflation equals the growth rate of nominal government liabilities, $\pi = g_N$.
4. Markets clear.
   (a) Money market:
   $$\int m(q(x,\epsilon))\, d\Phi(x,\epsilon) = M.$$
   (b) The real interest rate is equal to r^* as given by condition AMC.
   (c) Formal sector goods market:
   $$\int \left[1 - \hat{\jmath}^n\left(q(x,\epsilon)\right)\right] q(x,\epsilon)\, d\Phi(x,\epsilon) + \frac{G}{\mu_f} = \nu^F \int_{\hat{z}^n}^{\bar{z}} F\left(z, \ell_f(z)\right)\, d\Psi(z)$$
   (d) Informal sector goods market:
   $$\int \hat{\jmath}^n\left(q(x,\epsilon)\right) q(x,\epsilon)\, d\Phi(x,\epsilon) = \nu^F \int_{\underline{z}}^{\hat{z}^n} F\left(z, \ell_n(z)\right)\, d\Psi(z)$$

   By Walras' law, if all other markets clear, the labor market clears.
5. The household distribution is the fixed point of its law of motion and $\mathcal{T}$ is consistent with household's optimal behavior.

This definition of stationary equilibrium allows for general government policies: a policy mix is described as including six variables, (B, M, i, τ, G, g_N). The government has three degrees of freedom because its choices must satisfy the law of motion for government liabilities, inflation pins down the nominal rate (to match the exogenous real rate), and the nominal interest rate determines the demand for money.

3 Calibration

We calibrate the model for Peru, a representative developing country with a large informal sector. The calibration follows a multi-pronged approach. A first subset of parameters values are directly estimated from the data or set to standard values. The remaining parameters are chosen so that the model matches a series of targeted moments.

Functional forms

We use a Constant Relative Risk-Aversion utility function:

$$u(q,\ell)=\frac{q^{1-\sigma}-1}{1-\sigma}-\psi\frac{\ell^{1+\xi}}{1+\xi}.$$

The tax avoidance cost function $\chi_j(p)$ takes the following functional form:

$$\chi_j(p)=\chi_0(j)\,p^{\chi_1},\quad \chi_0(j)=\underline{\chi_0}^{1-j}\,\overline{\chi_0}^{j},\quad \overline{\chi_0}>\underline{\chi_0},\quad \chi_1>1.$$

For a given variety, the tax avoidance cost function is parameterized by two objects: $\chi_0(j)$ and χ_1. The first one determines the scale of enforcement costs, while the second determines how they grow with the size of the transaction. Because tax avoidance is easier for varieties with low j, we parameterize $\chi_0(j)$ with two parameters: $\underline{\chi_0}$ determines $\chi_0(j)$ for variety $j=0$ and $\overline{\chi_0}$ determines it for variety $j=1$.

For the financial services cost we follow Dotsey and Ireland (1996), Erosa and Ventura (2002), and Koreshkova (2006):

$$\gamma(j)=\gamma_0\left(\frac{1-j}{j}\right)^{\gamma_1}$$

The distribution of firm productivity $\Psi(z)$ is lognormal with parameters μ_z and σ_z. In turn, the process for idiosyncratic labor productivity is given by:

$$\ln(\epsilon_t)=\rho_\upsilon\ln(\epsilon_{t-1})+\upsilon_t,\quad \upsilon_t\sim N\left(0,\sigma_\upsilon^2\right)$$

We discretize this process following Tauchen (1986), which implies we only need to pick values for ρ_υ and σ_υ.

External calibration

Table 1 summarizes our choices with respect to the externally calibrated parameters. A model period corresponds to a quarter.[10] We set $\sigma=1$ (log preferences) and set the discount factor β so that the yearly discount factor equals 0.92 (Buera et al., 2011). The inverse elasticity of labor supply is set to 2.5, a value within the common range used for this elasticity. The degree of decreasing returns to scale in the production function, α, is set to 0.67, in line with the literature on firm dynamics in developing economies (Arellano et al., 2012). We equate both

[10] The length of the period in our model is important because with a cash-in-advance constraint the longer the period the lower the money velocity and the higher the seigniorage revenues from inflation. We check that our choice of a quarterly period results in an amount of seigniorage revenues in line with the data.

Table 1
Externally calibrated parameters

Parameter	Description	Value	Source
Households			
β	Discount factor	$0.92^{1/4}$	Buera et al. (2011)
σ	Inverse IES	1.0	
ξ	Inverse elasticity of labor supply	2.5	Elasticity of 0.4
Production Technology			
α	Decreasing returns	0.67	Arellano et al. (2012)
Government policy			
τ	Consumption tax	18%	VAT 16% + IPM 2%
π	Inflation	4%	CPI inflation
Other			
r^*	Real interest rate	3.2%	Deposit rate
t	Time period	3 months	

entry costs κ_e and average firm productivity (by setting μ_z) to 1 as a normalization.[11]

The real interest rate is set to 3.2% based on the real deposit rate, and our benchmark policy mix has a 18% consumption tax rate (corresponding to a 16% value added tax and a 2% municipal tax) and a 4% inflation rate.

Internal calibration

The internal calibration, summarized in Table 2, consists of choosing the values of 11 parameters to minimize the distance between 11 model-implied moments and their data counterparts.[12] Although each individual moment is jointly determined by all 11 parameters, when discussing calibration it is useful to relate each moment to the parameters that are most relevant to it. The disutility of labor parameter ψ determines labor supply, and we choose its value to normalize the steady-state aggregate labor supply to 1. We use ρ_v and σ_v to target the (yearly) autocorrelation and the dispersion of log earnings. The dispersion of firm productivity, σ_z, is chosen to match the dispersion of log firm size, measured by employment. The three values are 0.93, 0.42 and 0.16, respectively.

Parameters regarding the cost of formalization and of enforcement, κ_f, $\underline{\chi_0}$, $\overline{\chi_0}$ and χ_1, pin down the size of the informal sector, the expenditure on evasion costs, and how these costs evolve with j and q. Through their role in determining the equilibrium price schedules, they are also key to generating both the differential demand for informal consumption across expenditure levels, and how responsive are prices (μ_f, μ_n) to changes in government policy (such as τ). These 4 parameters are set jointly to target the following moments: *i*) the aggregate share of informality in the Peruvian economy; *ii*) the slope of the informality Engel curve; *iii*)

[11] Entry costs essentially determine average firm size: the mass of firms adjusts with entry costs such that $\kappa_e \times \nu^F$ is always close to $1 - \alpha$.

[12] See section C of the Appendix for details on the computation of our targets.

the passthrough of an increase in the consumption tax to the price of formal goods; and *iv*) the informality share of the first variety ($j = 0$). For the first moment, we use the average non-agricultural self-employment rate (50%), a common proxy for informal employment in the literature. For *ii*) and *iii*) we borrow from Bachas et al. (2024a). They estimate that the 5th percentile purchases about 70% of their consumption from the informal sector while the 95th percentile purchases only 35%. Accordingly, we target a slope of 0.35 for this relationship in the model. Unfortunately, there is no readily available estimate of the passthrough of consumption taxes into prices for Peru. Instead, we require that the model replicates the quasi-experimental evidence that Bachas et al. (2024a) document for Mexico, a country that shares several features regarding labor markets, taxes and the informal economy with Peru. Exploiting geographical variation in the exposure to a value-added tax (VAT) reform, they find a 75% passthrough of consumption taxes to prices in formal stores, and a 16% passthrough to prices in informal stores. We compute these moments in our model and use their estimate of the formal passthrough as a calibration target. Finally, we target an informality share of 100% (fully informal) for the first variety.

The parameters of the financial services cost function, γ_0 and γ_1, determine the relative cost of financial services and the responsiveness of money demand to changes in inflation. As Erosa and Ventura (2002), we set these values to match the average velocity of money and its semi-elasticity with respect to the nominal interest rate, which we estimate following Dotsey and Ireland (1996). Because we care about seigniorage as a source of revenue we focus on the monetary aggregate M0 (using M1 would lead us to overestimate seigniorage).[13]

Finally, we choose $\overline{D}$ to match a debt-to-GDP ratio of 33%.[14] Government revenues in Peru during our period were 19.38% of GDP. We do not target this directly because our model only has consumption and inflation taxes. Instead, we compute government expenditures endogenously, given the benchmark government policy, and keep them constant throughout our quantitative exercises.

Calibration results

Table 3 shows that the model does a good job at matching the targeted empirical moments. Figure 4a plots the informality Engel curve implied by the calibration, of which the slope was targeted. In the model, formal consumption increases from 15% of expenditure in the poorest decile to almost 50% for the richest decile. This is lower than the formal consumption in Bachas et al. (2024a). The level of the Informality Engel Curve is pinned down by the informal labor share, which also determines the size of the informal sector. Elgin et al. (2021) estimate an informal share of output for Peru of 53%, close to the one implied by our calibration

[13] To the extent that some components of M1 pay nominal interest rates that do not fully adjust with inflation, this choice implies that our results for the welfare cost of inflation are conservative.

[14] We consider public debt in terms of total liabilities of the government, including government bonds and money, defined as $D = B/(1+i) + M$. This is the most consistent with modern central banking: D is the total amount of public debt issued by the fiscal authority and open market operations change the composition of the outstanding liabilities between B and M, but not its total.

Table 2
Internally calibrated parameters

Parameter	Description	Value	Target
Households			
ψ	Disutility of labor	1.88	Unitary labor supply
ρ_ϵ	Persistence labor productivity	0.93	Autocorrelation of log income
σ_ϵ	Std dev of labor productivity shocks	0.42	St. dev. of log income
Production Technology			
σ_z	Firm productivity dispersion	0.16	St. dev. of log firm size
κ_f	Formalization cost	1.35	Informal sector share
Tax avoidance costs			
$\underline{\chi_0}$	Tax avoidance cost scale $j = 0$	0.18	Informality share of variety $j = 0$
$\overline{\chi_0}$	Tax avoidance cost scale $j = 1$	0.44	Slope informality Engel curve
χ_1	Tax avoidance cost exponent	1.10	Consumption tax passthrough
Financial services costs			
γ_0	Financial services cost scale	0.27	Money velocity
γ_1	Financial services cost slope	1.55	Semi-elasticity of income velocity
Government policy			
$\overline{D}$	Public debt	1.54	Debt-to-GDP ratio

Table 3
Model fit

Moment	Target	Model
Money-to-GDP ratio	0.21	0.19
Semi-elasticity of velocity	8.00	8.01
Debt-to-GDP ratio	0.33	0.33
Autocorrelation of log income	0.91	0.92
Std. dev. of log income	1.60	1.56
Std. dev. of log firm size	0.66	0.68
Slope of Engel curve for informality	0.35	0.35
Consumption tax passthrough	0.75	0.84
Size of the informal sector (labor)	0.50	0.51
Share of informality of variety $j = 0$	1.00	1.00

at 57%.

Formalization costs are 35% higher than entry costs, results quantitatively in line with the estimates of Ulyssea (2018) for Brazil. The model predicts that approximately 83% of firms operate informally. Although a direct counterpart for this measure is not readily available, 94.2% of firms in Peru are microenterprises, and 99% are microenterprises plus small firms, which are predominantly informal (Quispe et al., 2024). The model delivers a satisfactory match to the targeted passthrough of consumption taxes to formal prices and, importantly, it also matches Bachas et al.'s estimate of the informal passthrough at 16%, even though this moment was not directly targeted. In addition, the calibrated values for the cost of financial services imply that 85% of transactions in the model economy are made in cash, close to the 80% found by Aurazo and Vega (2021).

The model accurately matches the relative importance of consumption taxes and seigniorage in the government's budget. Seigniorage revenues add up to 0.77% of GDP, which amounts to approximately 10% of total government revenues in the model. When applying the methodology of Aisen and Veiga (2008) to compute seigniorage using Peruvian data in the time window we study, we compute seigniorage values of 0.5% and 1.6% of GDP, depending on the methodology. Revenue from consumption (VAT) taxes is 6.3% in the model, while it is 7.6% in the data.[15] With its revenues the government can finance government expenditures totaling 6.5% of GDP after paying interest expenditures on its debt.

4 Informality and fiscal progressivity

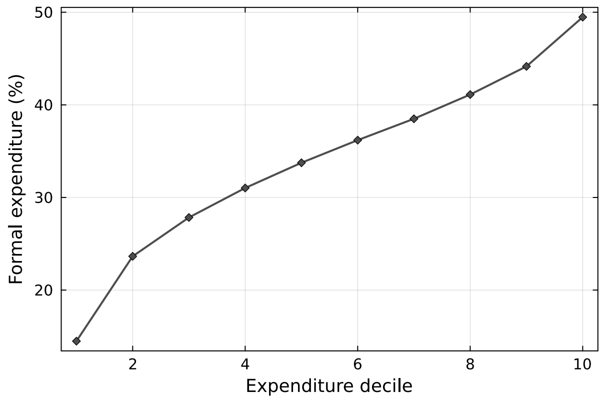
In this section we discuss the distributional consequences of the non-homothetic demand for informal consumption. We first show how, under the benchmark policy, poorer households benefit more from the lower prices offered by informal producers. We then document how households' exposure to each tax – and their overall effective tax rates – vary substantially across the expenditure distribution, shaping the overall progressivity of fiscal policy. Next, we examine the aggregate and distributional effects of revenue-neutral changes on inflation.

Distributional outcomes under the benchmark policy

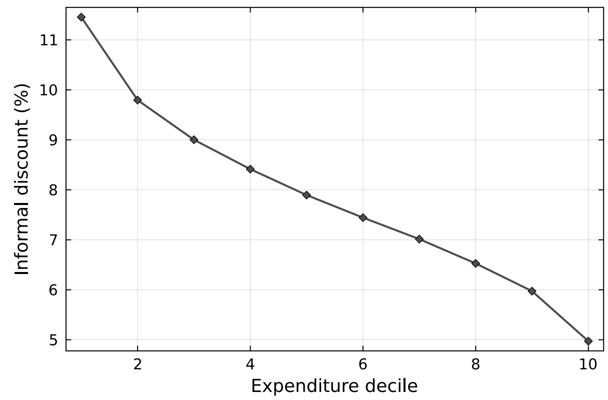
An important implication of the model is that, because informal firms avoid the burden of taxation and the payment of formalization costs, they are able to sell low quality goods at cheaper prices than the formal sector. Figure 4b plots, for each expenditure decile, the complement of the ratio between household's expenditure and the expenditure necessary to purchase their consumption bundle entirely from the formal sector, which we label as the 'informal discount'. Hence, the informal discount captures the effect that the informal sector has on households' purchasing power. The average informal discount implied by the calibration is substantial: on average, households spend 7% less than the expenditure required to purchase an *identical* consumption bundle solely from the formal sector. Moreover, poorer households achieve substantially higher informal discounts due to their higher demand for informal goods. Indeed, the poorest decile averages an informal discount of 11%, whereas the informal discount for the highest expenditure decile is 5%. These numbers are in the ballpark of findings by Bachas et al. (2024a), who show that in Peru prices for similar goods in formal retailers are on average 15% higher than in informal ones.

Our findings that the informal sector's ability to bypass costly taxes and regulation translates into cheaper prices for goods disproportionately bought by poorer households parallels

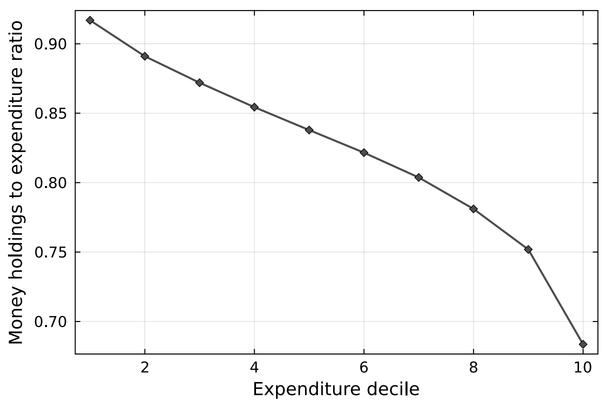
[15] Banco Central de Reserva del Perú, retrieved from here.



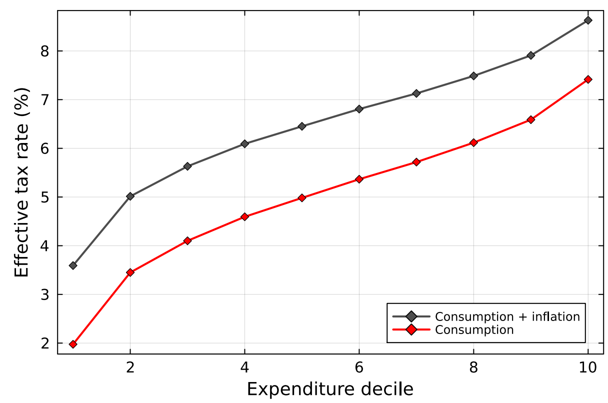
(a) Formality Engel curve



(b) Informal discount



(c) Money holdings ratio



(d) Effective tax rates

Figure 4
Heterogeneity by expenditure decile

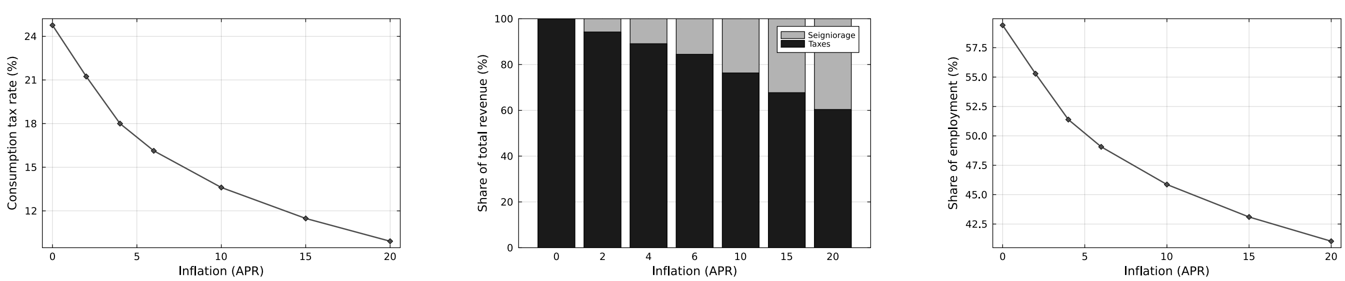
findings highlighting its role as a flexible source of employment opportunities that ameliorates the negative employment effects of said taxes and regulations, especially so for low skilled workers (La Porta and Shleifer, 2014; Almeida and Carneiro, 2012; Ulyssea, 2010). The heterogeneous incidence of informal activity might have important consequences for the large body of work showing how informality reduces allocative efficiency, with adverse consequences for productivity and output (see Ulyssea (2020) for a review). A common result in this literature is that stricter enforcement of informality can lead to significant productivity gains, but these analyses typically are mute about potential distributional consequences – which, as our results suggest, may be substantial.

As in Bachas et al. (2024a), the *de facto* exemption of informal consumption from consumption taxes implies that the latter are strongly progressive – the poorest decile pays an effective consumption tax of 2%, whereas the wealthiest decile pays more than 3 times as much (see Figure 4d). What is different in our framework is that the costly financial services and the greater reliance on informal consumption by poorer households result in increased exposure to the inflation tax due to higher cash holdings (see Figure 4c). As a result, the progressivity achieved by consumption taxes is partially undone by the regressiveness of inflation

(Erosa and Ventura, 2002). Although the difference between the effective consumption taxes paid by the richest and poorest deciles is 5.4%, this number drops to 5% once we account for inflation; this is equivalent to increasing the relative tax rate paid by the poorest decile from 27 to 42%. This heterogeneous exposure to different taxes across the expenditure distribution generates different preferences for revenue-equivalent tax combinations, something we explore in the next section.

Effects of revenue-neutral fiscal reform

We study the consequences of revenue-neutral policy changes by varying inflation and adjusting the consumption tax rate to keep total government revenue unchanged at the benchmark policy level; details on the algorithm we use can be found in Section C of the Appendix. Figure 5 displays the results. Higher inflation allows for a reduction in the consumption tax rate by partially substituting consumption tax revenues with seigniorage. As shown in Figure 5a, different levels of long-run inflation generate substantial changes in consumption taxes, a result driven by the effectiveness of seigniorage as a source of revenue implied by the calibration. Figure 5b details the composition of government revenues between tax and seigniorage revenues. Increasing inflation permanently by 2% leads to a decrease of the consumption tax rate by 1.9 percentage points, and the share of seigniorage in total government revenue increases from 11% to 15.5%. Increasing inflation discourages cash use and encourages households to rely more on credit, reducing their informal purchases (see Figure 5c). As demand shifts toward formal-sector goods, their prices rise, and the productivity threshold at which firms find it optimal to formalize lowers, increasing formality. As with consumption taxes, the informal sector also responds strongly to permanent changes in inflation: increasing inflation by 2% reduces the employment share of the informal sector by 2.3 percentage points (or a 4.5% change).

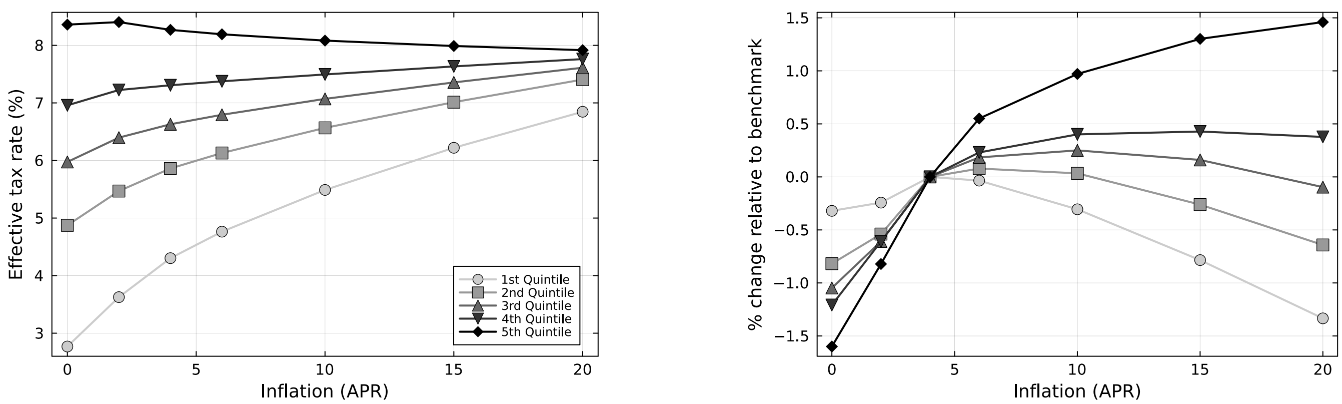


(a) Consumption tax rate (b) Revenue composition (c) Size of the informal sector

Note: Inflation measured in annualized percentage rate (APR). For 0% inflation, government revenue is 2.8% below the benchmark level because of the limits imposed by the Laffer curve of taxation.

Figure 5
Effects of policy change on aggregate variables

Figure 6a shows how a shift towards greater inflation and lower consumption taxes significantly reduces the differences in effective tax rates across expenditure groups, thus reducing the overall progressivity of the tax system. Under the benchmark policy, the effective tax rate of the top quintile is 8.3%, 92% higher than the 4.3% rate faced by the bottom quintile. This gap narrows to 47% under 10% inflation, and to just 16% under 20% inflation (with the top quintile paying 7.9% and the bottom quintile paying 6.8%). Figure 6b illustrates how this reduction in fiscal progressivity amplifies consumption inequality. At 10% inflation, the consumption quality of the top quintile is 1% higher than in the benchmark, while the bottom quintile's is 0.3% lower.



(a) Effective tax rates

(b) Consumption quality

Notes: Inflation measured in annualized percentage rate (APR). For 0% inflation, government revenue is 2.8% below the benchmark level because of the limits imposed by the Laffer curve of taxation.

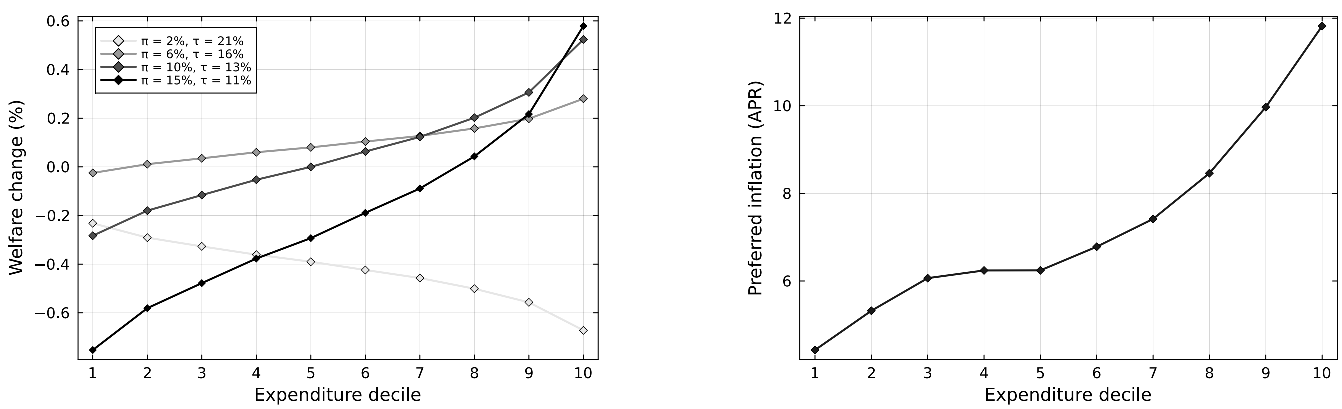
Figure 6
Effects of policy change across the expenditure distribution

What do these substantial changes in fiscal progressivity imply for welfare? It has been well established since Nicolini (1998) that the Friedman Rule is no longer optimal in economies with a sizable informal sector. Building on this insight, the quantitative analyses of Koreshkova (2006) and Aruoba (2021) show that, in representative-agent models, optimal inflation rates increase significantly with the size of the informal sector. Our model incorporates the same core mechanisms, but allows us to evaluate the welfare impacts across the expenditure distribution – because households differ in their reliance on informal consumption, they also differ in their preferred level of inflation. To do so, we compute the consumption-equivalent variation associated with moving away from the benchmark policy. This measure indicates the permanent change in consumption quality under the benchmark policy that would make a household as well-off as after the policy change.[16] The distribution of the consumption-equivalent welfare effects is shown in Figure 7a. Households unanimously prefer the benchmark policy to one with 2% inflation. This result exemplifies the relevance of the distortionary effects of formal taxation that undermine the optimality of the Friedman Rule. Take the case of households in the bottom quintile. As displayed in Figure 6a, under the benchmark policy they face a substantially higher effective tax rate than under the policy of 2% inflation. However, while this

[16]This measure is conceptually similar to consumption-equivalent variation, commonly used to evaluate the welfare effects of policy. However, in our setting, utility depends on the quality of the consumption bundle rather than a quantity. In the formal sector, where prices are linear, a change in quality maps directly to a change in expenditure. This equivalence breaks down in the informal sector due to the non-linear relationship between pricing and quality.

policy change would decrease their effective tax rate, it would have such large impacts on the aggregate productivity of the economy that the resulting reduction in household income would make these households worse off. Consistent with this, raising inflation to 6% improves welfare for most households: although the bottom 10% experience a loss in welfare, all others benefit, with the top decile gaining the most (a 0.25% consumption-equivalent increase). A policy of 10% inflation and 13% consumption tax rate benefits the upper half of the distribution at the expense of the lower half. The bottom quintile loses about 0.25%, while the top quintile gains approximately 0.4%. At 15% inflation, with the consumption tax reduced further to 11%, the policy becomes welfare-reducing for a clear majority of households – yet the top 25% continue to benefit.

Figure 7b presents the same information from a different perspective, showing each household's preferred inflation rate across the expenditure distribution. This is computed by finding the policy mix that maximizes a household's expected lifetime discounted utility given its state variables. There is very clear and significant dispersion in desired inflation rates. The bottom decile of households would like inflation to be below 5%, while the top decile wants an inflation rate around 12%. The median household's preferred inflation rate is 6.5%.[17]



(a) Welfare effects of policy change

(b) Preferred inflation

Notes: Preferred inflation measured in annualized percentage rate (APR).

Figure 7
Policy change and welfare

Removing the costs from the credit payment technology

Our results on the heterogeneity in money holdings and in preferences for inflation is a consequence of two different model mechanisms: the non-homothetic demand for informal consumption and economies of scale in the credit payment technology (Erosa and Ventura, 2002). To illustrate the quantitative relevance of the mechanism tied to informal consumption, we shut down the costs of using credit payments in the formal sector by setting $\gamma_0 = 0$, and recompute the welfare effects of different policy mixes.

[17]Section C in the Appendix illustrates how earnings inequality amplifies these distributional considerations.

With costless access to credit payments, households will use credit in all their formal purchases, and will only use cash to finance informal consumption. Table 4 compares distributional outcomes between the benchmark model and the counterfactual with costless credit payments. Notably, effective tax rates in the counterfactual display the same degree of progressivity as in the benchmark. Moreover, the heterogeneity in the welfare changes of moving from 4 to 10% inflation remain substantial (the interquartile range is 0.32 p.p. with financial cost services and 0.24 p.p. in the counterfactual without it); the levels differ from the benchmark because without costly credit payments inflation is less distortionary and, as a consequence, optimal inflation is higher.

These results imply that the non-homothetic demand for informal consumption is enough to generate sizeable distributional implications of inflation in developing countries. Moreover, they suggest that as long as the informal sector is excluded from the credit payments system, these distributional concerns might persist even in the face of improvements to access to banking, credit cards, and financial innovations such as digital payment systems.

Table 4
Removing the cost of credit payments

		Benchmark	Free credit
Effective tax rate (%)	P25	5.7	5.8
	P50	6.7	6.8
	P75	7.5	7.6
Welfare change (%)	P25	-0.21	0.58
$\pi = 4\% \rightarrow \pi = 10\%$	P50	-0.04	0.71
	P75	0.11	0.82

Notes: Column "Free credit" corresponds to the case with $\gamma_0 = 0$.

Varying the size of the informal sector

The literature on Ramsey optimal taxation with informality and representative agents can rationalize, both theoretically and quantitatively, the positive relationship between informality and inflation in the data (Koreshkova, 2006; Aruoba, 2021). In this section we contribute to this literature by using the calibrated model to study how the distributional effects of seigniorage vary as we change the size of the informal sector, which we achieve by changing formalization costs κ_f.[18] Table 5 displays the informal discount and the welfare changes associated with increases in inflation for different expenditure quartiles, for both the benchmark model and two counterfactual economies (high and low informality). In the "low informality" economy the size of the informal sector, measured by its share of labor demand, equals 25%, a level comparable to Chile, whereas this number is 70% in the "high informality" counterfactual, similar to

[18] We also adjust government expenditure so that the government maintains a balanced budget across counterfactuals.

Angola. Changing the size of the informal sector significantly amplifies the distributional considerations associated with inflation. Indeed, moving from the low to the high informality counterfactual increases the interquartile range of the informal discount from 1.6 p.p. to 4.7 p.p. Similarly, the interquartile range of the welfare changes associated with higher inflation also increase from 0.2 p.p. to 0.46 p.p.

Table 5
The role of informality

		Benchmark	Low informality	High informality
Informal discount (%)	P25	10.1	3.2	19.2
	P50	8.3	2.2	16.3
	P75	7.1	1.6	14.5
Welfare change (%)	P25	-0.21	-0.38	-0.30
$\pi = 4\% \rightarrow \pi = 10\%$	P50	-0.04	-0.28	-0.10
	P75	0.11	-0.18	0.16

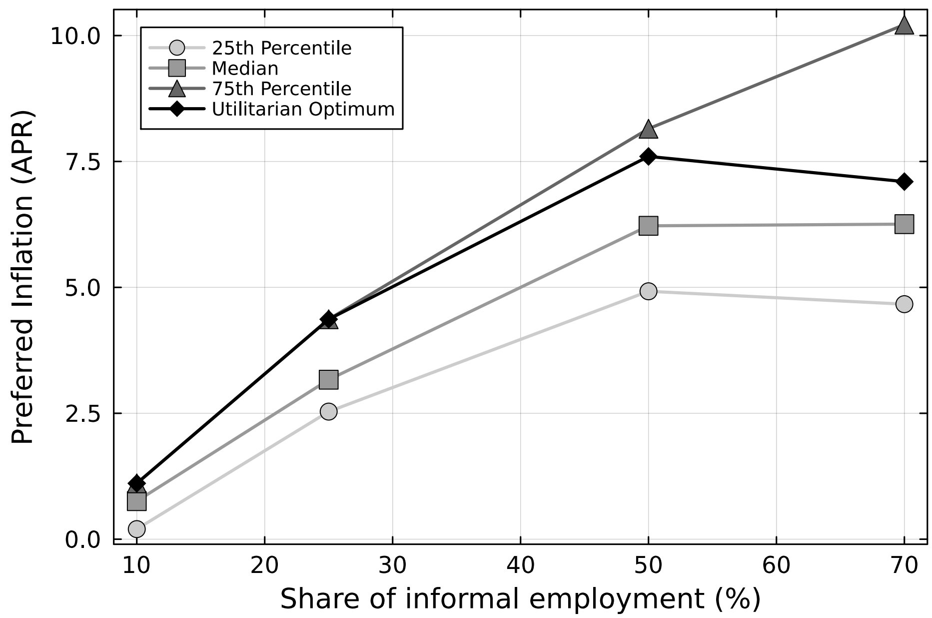
Notes: “Low informality” and “High informality” correspond to economies with an informal sector size of 25% and 70% respective. See text for details.

Figure 8 further illustrates the increase in the distributional effects of inflation as informality increases. It plots both optimal inflation (from an utilitarian perspective) as well as the desired inflation for households along the wealth distribution in economies with different degrees of informality. Consistent with the findings of Koreshkova (2006) and Aruoba (2021), optimal inflation is increasing in the size of the informal sector. Moreover, we observe how the dispersion between the desired levels of inflation increases substantially with the size of the informal economy, highlighting how inflation is more regressive. Overall, these results indicate that the relevance of the distributional concerns studied in this paper is even larger for countries at earlier stages of the development spectrum, as they feature larger informal economies.

5 Conclusion

This paper examines how the coexistence of large informal sectors and limited fiscal capacity in developing countries shapes the distributional effects of indirect taxation instruments. In it, we have developed a dynamic general equilibrium where formalization decisions, sectoral prices, and household choices over where to shop and how to pay are jointly determined. By doing so, the framework highlights that the informal sector’s ability to evade oversight generates sizable “informal discounts” that disproportionately benefit the poor, but also that the cash-intensive nature of informal activity heightens their exposure to inflation.

Our quantitative analysis reveals that the trade-off between consumption taxation and seigniorage has substantial distributional implications. The *de facto* exemption of informal consumption from consumption taxes makes them progressive in our setting, but implies that



Notes: The share of informal employment is measured under the benchmark policy of 4% inflation and 18% consumption tax rate. Preferred inflation measured in annualized percentage rate (APR).

Figure 8
Preferred inflation in economies with different level of informality

inflation partly undoes this progressivity by increasing money holdings on the left tail of the expenditure distribution. Revenue-neutral reforms that substitute inflation for consumption taxes benefit richer households at the expense of poorer ones, and the degree of disagreement grows with the prevalence of informality. Acknowledging these distributional dimensions is relevant when considering the use of seigniorage as a source of revenue in developing countries.

These findings point to several promising avenues for future research. First, while our model focuses on inflation and consumption taxes, it could be extended to study other type of policies related to financial inclusion, payment systems, and the enforcement of taxation. Secondly, another important source of distributional impacts of public finance decisions in economies with a large informal sector is the labor market which is also strongly segmented. We have chosen to focus on the consumption side in our analysis, but future work on the labor market side would complement the results presented above.

References

Ari Aisen and Francisco José Veiga. The political economy of seigniorage. *Journal of Development Economics*, 87(1):29–50, 2008. ISSN 03043878. doi: 10.1016/j.jdeveco.2007.12.006.

Rita Almeida and Pedro Carneiro. Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics*, 4(3):64–89, 2012. ISSN 19457790. doi: 10.1257/app.4.3.64.

Cristina Arellano, Yan Bai, and Jing Zhang. Firm dynamics and financial development. *Journal of Monetary Economics*, 59(6):533–549, 2012. ISSN 03043932. doi: 10.1016/j.jmoneco.2012.06.006. URL http://dx.doi.org/10.1016/j.jmoneco.2012.06.006.

S. Borağan Aruoba. Institutions, tax evasion, and optimal policy. *Journal of Monetary Economics*, 118:212–229, 2021. ISSN 03043932. doi: 10.1016/j.jmoneco.2020.10.003. URL https://doi.org/10.1016/j.jmoneco.2020.10.003.

Zareh Asatryan and David Gomtsyan. The Incidence of VAT Evasion. *CESifo Working Paper Series*, 2020.

Jose Aurazo and Milton Vega. Why people use digital payments: Evidence from micro data in Peru. *Latin American Journal of Central Banking*, 2(4):100044, 2021. ISSN 26661438. doi: 10.1016/j.latcb.2021.100044. URL https://doi.org/10.1016/j.latcb.2021.100044.

Pierre Bachas, Lucie Gadenne, and Anders Jensen. Informality, Consumption Taxes, and Redistribution. *Review of Economic Studies*, 91(5):2604–2634, 2024a. ISSN 1467937X. doi: 10.1093/restud/rdad095.

Pierre Bachas, Anders Jensen, and Lucie Gadenne. Tax Equity in Low- and Middle-Income Countries. *Journal of Economic Perspectives*, 38(1):55–80, 2024b.

Timothy Besley and Torsten Persson. Why Do Developing Countries Tax So Little? *Journal of Economic Perspectives*, 28(4):99–120, 2014. ISSN 08953309. doi: 10.1257/jep.28.4.99. URL http://dx.doi.org/10.1257/jep.28.4.99.

Francisco J Buera, Joseph P Kaboski, and Yongseok Shin. Finance and Development: A Tale of Two Sectors. *The American Economic Review*, 2011. doi: 10.1257/aer.101.5.1964. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1964.

Jordi Caballé and Judith Panadés. Inflation, tax evasion, and the distribution of consumption. *Journal of Macroeconomics*, 26(4):567–595, 2004. ISSN 01640704. doi: 10.1016/j.jmacro.2003.06.001.

Shutao Cao, Césaire A. Meh, José Víctor Ríos-Rull, and Yaz Terajima. The welfare cost of inflation revisited: The role of financial innovation and household heterogeneity. *Journal of*

Monetary Economics, 118:366–380, 2021. ISSN 03043932. doi: 10.1016/j.jmoneco.2020.11.004. URL https://doi.org/10.1016/j.jmoneco.2020.11.004.

Alberto Chong and Mark Gradstein. Inequality and informality. *Journal of Public Economics*, 91(1-2):159–179, 2007. ISSN 00472727. doi: 10.1016/j.jpubeco.2006.08.001.

Fernando Cirelli. Bank-Dependent Households and The Unequal Costs of Inflation. *Working paper*, 2022. URL www.fernandocirelli.com.

Fernando Cirelli, Emilio Espino, and Juan M. Sánchez. Designing unemployment insurance for developing countries. *Journal of Development Economics*, 148(September 2020):102565, 2021. ISSN 03043878. doi: 10.1016/j.jdeveco.2020.102565. URL https://doi.org/10.1016/j.jdeveco.2020.102565.

Pablo N. D'Erasmo and Hernan J. Moscoso Boedo. Financial structure, informality and development. *Journal of Monetary Economics*, 59(3):286–302, 2012. ISSN 03043932. doi: 10.1016/j.jmoneco.2012.03.003. URL http://dx.doi.org/10.1016/j.jmoneco.2012.03.003.

Michael Dotsey and Peter Ireland. The welfare cost of inflation in general equilibrium. *Journal of Monetary Economics*, 37(1):29–47, 1996. ISSN 03043932. doi: 10.1016/0304-3932(95)01239-7.

Ceyhun Elgin, M. Ayhan Kose, Franziska Ohnsorge, and Shu Yu. Understanding Informality. *CAMA Working Paper*, 76, 2021.

Andrés Erosa and Gustavo Ventura. On inflation as a regressive consumption tax. *Journal of Monetary Economics*, 49(4):761–795, 2002. ISSN 03043932. doi: 10.1016/S0304-3932(02)00115-0.

Daniel Jaar. Self-employment as Self-insurance. *SSRN*, 2024.

Clement Joubert. Pension design with a large informal labor market: Evidence from Chile. *International Economic Review*, 56(2):673–694, may 2015. ISSN 14682354. doi: 10.1111/IERE.12118.

Tatyana A. Koreshkova. A quantitative analysis of inflation as a tax on the underground economy. *Journal of Monetary Economics*, 53(4):773–796, 2006. ISSN 03043932. doi: 10.1016/j.jmoneco.2005.02.009.

Rafael La Porta and Andrei Shleifer. Informality and Development. *The Journal of Economic Perspectives*, 28(3):109–126, 2014. ISSN 19447965. doi: 10.1257/jep.28.3.109. URL http://dx.doi.org/10.1257/jep.28.3.109.

Costas Meghir, Renata Narita, and Jean Marc Robin. Wages and informality in developing countries. *American Economic Review*, 105(4):1509–1546, 2015. ISSN 00028282. doi: 10.1257/aer.20121110.

Abhinav Narayanan, Gita Gopinath, Prachi Mishra, and Gabriel Chodorow-Reich. Cash and the Economy: Evidence from India's Demonetization. *The Quarterly Journal of Economics*, pages 57–103, 2020. doi: 10.1093/qje/qjz027.Advance.

Juan Pablo Nicolini. Tax evasion and the optimal inflation tax. *Journal of Development Economics*, 55(1):215–232, 1998. ISSN 03043878. doi: 10.1016/s0304-3878(97)00063-1.

Pandia Kelly Quispe, Tantalean Sergio Saldaña, Echevarria Raquel Rengifo, Seminario Marlon Broncano, Quintana Nelly Rivera, and Casanova Huiman Susana. Las Mipymes en cifras 2022. *Ministerio de la Producción*, page 185, 2024. URL file:///C:/Users/usuario/Downloads/3Libro{_}Las{_}Mipyme{_}en{_}Cifras{_}2022{_}OGEIEE{_}web{_}18.01.2024{_}f{_}1{_}1.pdf{%}0Ahttps://ogeiee.produce.gob.pe/index.php/en/shortcode/oee-documentos-publicaciones/publicaciones-anuales/item/1170-las-mipyme-en-cifras-2022.

George Tauchen. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181, jan 1986. ISSN 01651765. doi: 10.1016/0165-1765(86)90168-0. URL https://linkinghub.elsevier.com/retrieve/pii/0165176586901680.

Gabriel Ulyssea. Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics*, 91(1):87–99, 2010. ISSN 03043878. doi: 10.1016/j.jdeveco.2009.07.001. URL http://dx.doi.org/10.1016/j.jdeveco.2009.07.001.

Gabriel Ulyssea. Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review*, 108(8):2015–2047, 2018. ISSN 00028282. doi: 10.1257/aer.20141745.

Gabriel Ulyssea. Informality: Causes and Consequences for Development. 2020. doi: 10.1146/annurev-economics. URL https://doi.org/10.1146/annurev-economics-.

A Market clearing

Let $h \in [0,1]$ denote a household's index and $k \in [0, \nu^F]$ denote a firm's index, where ν^F is the mass of entrants. Denote the firm's decision of which variety to produce as $v_k \in [0,1]$, the sectoral decision as $g_k \in \{f, n\}$, and the amount of goods of quality q produced as $s_k(q) \in \mathbb{R}_0^+$. Let $A \subseteq \mathbb{R}^+$ be a Borel subset of the space of qualities and $J \subseteq [0,1]$ be a Borel subset of the space of varieties.

Equilibrium in the formal sector, where both households and the government demand goods, is given by:

$$\int_0^1 \int_J \mathbb{I}\{q_{hjft} \in A\} dj dh + \int_J \int_A g_{jt}(q) dq dj = \int_0^{\nu^F} \left(\mathbb{I}\{v_{kt} \in J\} \mathbb{I}\{g_{kt} = f\} \int_A s_{kt}(q)\, dq \right) dk, \forall A \subseteq \mathbb{R}^+, J \subseteq [0,1].$$

In the informal sector, only households are on the demand side:

$$\int_0^1 \int_J \mathbb{I}\{q_{hjnt} \in A\}\, dj\, dh = \int_0^{\nu^F} \left(\mathbb{I}\{v_{kt} \in J\} \mathbb{I}\{g_{kt} = n\} \int_A s_{kt}(q)\, dq \right) dk,\ \forall A \subseteq \mathbb{R}^+,\ J \subseteq [0,1].$$

Labor market clearing is given by:

$$\int l_{kt}\, dk + \kappa_e \nu_t^F + \int \kappa_f \mathbb{I}\{g_{kt} = f\}\, dk + \bar{\chi}_t + \bar{\gamma}_t = \int \epsilon_{ht} l_{ht}\, dh + \left(\int b_{ht}\, dh - B_t \right) \left[1 - \left(\frac{1}{1 + i_t} \right) \right]$$

where $\bar{\chi}_t$ is the total amount of tax avoidance costs firms incur, and $\bar{\gamma}_t$ is the total amount of financial services costs. The last term on the right-hand side of the equation refers to the hiring of foreign labor as a counterpart to foreign asset holdings. The expressions for $\bar{\chi}_t$ and $\bar{\gamma}_t$ are:

$$\bar{\chi}_t = \int \int \chi_j \left(p_{jnt}(q_{hjnt}) \right)\, dj\, dh$$

$$\bar{\gamma}_t = \int \int \mathbb{I}\{j \notin \mathcal{J}_{ht}^m\}\, \gamma(j)\, dj\, dh$$

The money market clearing condition is given by:

$$\int m_{ht}\, dh = M_t$$

B Proofs

Proposition 1

The revenue maximization of a formal firm of variety j with productivity z and ℓ units of labor is given by:

$$R_{jf}(z, \ell) = \max_{s(q)} \int \left(\frac{p_{jf}(q)}{1 + \tau} \right) s(q) dq$$

$$s.t.: \quad \int q s(q)\, dq \leq F(z, \ell)$$

The first order condition of this problem with respect to $s(q)$ is:

$$\left(\frac{p_{jf}(q)}{1 + \tau} \right) - \mu_{jf} q = 0,$$

where μ_{jf} is the Lagrangian multiplier on the constraint.

The revenue maximization of an informal firm of variety j with productivity z and ℓ units of labor is given by:

$$
\begin{aligned}
R_{jn}(z,\ell) = & \max_{s(q)} \int \left[p_{jn}(q) - \chi_j\left(p_{jn}(q)\right)\right] s(q) dq \\
s.t.: & \quad \int qs(q)\, dq \leq F(z,\ell)
\end{aligned}
$$

The first order condition of this problem with respect to $s(q)$ is:

$$
\left[p_{jn}(q) - \chi_j\left(p_{jn}(q)\right)\right] - \mu_{jn} q = 0,
$$

where μ_{jn} is the Lagrangian multiplier on the constraint.

These two conditions correspond to equation (1). Plug them back into the revenue expressions to write:

$$
R_{jf}(z,\ell) = \mu_{jf} F(z,\ell), \quad R_{jn}(z,\ell) = \mu_{jn} F(z,\ell),
$$

implying that profit for a firm producing variety j in sector s with productivity z is given by:

$$
\Pi_{js}(z) = \max_{\ell} \mu_{js} F(z,\ell) - w\ell.
$$

Due to free entry across varieties, a firm with productivity z in sector s must be indifferent between all varieties j. This can only be the case if $\mu_{js} = \mu_s, \forall j \in [0,1]$.

Proposition 2

A firm wants to become formal iff

$$
\Pi_f(z) - \kappa_f \geq \Pi_n(z).
$$

Define the function $g : [\underline{z}, \overline{z}] \to \mathbb{R}$ as:

$$
g(z) \equiv \Pi_f(z) - \kappa_f - \Pi_n(z).
$$

Clearly, g is continuous.

Theorem 5 *If g is strictly increasing in z, then the threshold rule in Proposition 2 holds.*

Proof. Three different cases are possible, all leading to the threshold rule:

1. If $g(\underline{z}) < 0$ and $g(\overline{z}) > 0$, then by the intermediate value theorem $\exists \hat{z}^n \in (\underline{z}, \overline{z}) : g(\hat{z}^n) = 0$. Additionally, strict monotonicity implies $g(z) < 0, \forall z \in [\underline{z}, \hat{z}^n)$, and $g(z) > 0, \forall z \in (\hat{z}^n, \overline{z}]$;

2. If $g(\underline{z}) > 0$, strict monotonicity implies $g(z) > 0, \forall z \in [\underline{z}, \overline{z}]$, so for $\hat{z}^n = \underline{z}$ the threshold rule applies (and all firms formalize);

3. If $g(\overline{z}) < 0$, strict monotonicity implies $g(z) < 0, \forall z \in [\underline{z}, \overline{z}]$, so for $\hat{z}^n = \overline{z}$ the threshold rule applies (and all firms operate informally).

■

Theorem 6 *If $\mu_f > \mu_n$ and the production function F satisfies $F_z > 0$, $F_\ell > 0$, $F_{\ell z} > 0$, $F_{\ell\ell} < 0$ and $F(0, \ell) = 0$, then g is strictly increasing in z.*

Proof.

Define the function $\tilde{\Pi} : \mathbb{R}^+ \times [\underline{z}, \overline{z}] \to \mathbb{R}$:

$$\tilde{\Pi}(\mu, z) = \max_{\ell} \mu F(z, \ell) - w\ell.$$

and note that

$$\Pi_f(z) = \tilde{\Pi}(\mu_f, z), \quad \Pi_n(z) = \tilde{\Pi}(\mu_n, z).$$

Then:

$$g'(z) = \Pi_f'(z) - \Pi_n'(z) = \frac{\partial \tilde{\Pi}(\mu_f, z)}{\partial z} - \frac{\partial \tilde{\Pi}(\mu_n, z)}{\partial z} = \int_{\mu_n}^{\mu_f} \frac{\partial^2 \tilde{\Pi}(\tilde{\mu}, z)}{\partial z \partial \mu} \, d\tilde{\mu},$$

where the last equivalence uses the fundamental theorem of calculus. Note that:

$$\frac{\partial^2 \tilde{\Pi}(\mu, z)}{\partial z \partial \mu} > 0, \; \forall \mu > 0, z \in [\underline{z}, \overline{z}] \Rightarrow \int_{\mu_n}^{\mu_f} \frac{\partial^2 \tilde{\Pi}(\tilde{\mu}, z)}{\partial z \partial \mu} \, d\tilde{\mu} > 0 \Leftrightarrow g'(z) > 0.$$

We prove that the cross-derivative of $\tilde{\Pi}$ is indeed positive. Use the envelope theorem to get:

$$\frac{\partial \tilde{\Pi}(\mu, z)}{\partial \mu} = F\left(z, \ell^*(\mu, z)\right),$$

and further differentiate with respect to z:

$$\frac{\partial^2 \tilde{\Pi}(\mu, z)}{\partial z \partial \mu} = F_z\left(z, \ell^*(\mu, z)\right) + F_\ell\left(z, \ell^*(\mu, z)\right) \frac{\partial \ell^*(\mu, z)}{\partial z}.$$

All the terms are positive, hence, this derivative is positive and g is strictly increasing:

1. The first term is positive because we assumed $F_z > 0$.
2. The second term is positive because we assumed $F_\ell > 0$.
3. $\ell^*(\mu, z)$ is implicitly defined by the first order condition:

$$\mathcal{F}(\mu, z, l) \equiv \mu F_\ell(z, \ell) - w = 0.$$

Using the implicit function theorem:

$$\frac{\partial \ell^*(\mu, z)}{\partial z} = -\frac{\frac{\partial \mathcal{F}}{\partial z}}{\frac{\partial \mathcal{F}}{\partial \ell}} = -\frac{\mu F_{\ell z}(z, \ell)}{\mu F_{\ell\ell}(z, \ell)} = -\frac{F_{\ell z}(z, \ell)}{F_{\ell\ell}(z, \ell)} > 0$$

The inequality follows from $F_{\ell z} > 0$, $F_{\ell\ell} < 0$.

■

Whenever $\mu_f > \mu_n$ is not true, $\Pi_f(z) < \Pi_n(z), \forall z \in [\underline{z}, \overline{z}]$ and no firm chooses to be formal.

Proposition 3

Statement 1

Theorem 7 *If $\chi_j' > 0$ and $\chi_j'' > 0$, then $p_{jn}'(q) > 0$ and $p_{jn}''(q) > 0$, $\forall q \in (0, \bar{q}_{jn})$. $p_{jn}(q)$ is not defined for $q > \bar{q}_{jn}$, meaning that there is no supply of informal goods for those quality levels.*

Proof. Write the equilibrium price schedule for informal goods as:

$$p_{jn} - \chi_j\left(p_{jn}\right) = \mu_n q_{jn}.$$

If $\lim_{p\to\infty} \chi_j'(p) > 1$, the left-hand side of the expression above has a maximum at $\bar{p}_{jn}$ such that $\chi_j'(\bar{p}_{jn}) = 1$ and $\chi'(p) < 1, \forall p \in (0, \bar{p}_{jn})$. $\bar{q}_{jn}$ defined such that

$$\bar{p}_{jn} - \chi_j\left(\bar{p}_{jn}\right) = \mu_n \bar{q}_{jn},$$

is the highest quality for which there is a price that satisfies the equilibrium price schedule. The price schedule is not defined for $q > \bar{q}_{jn}$.

The first derivative of the price schedule is:

$$p_{jn}'(q) = \mu_n + \chi_j'\left(p_{jn}(q)\right)\ p_{jn}'(q) \implies p_{jn}'(q) = \frac{\mu_n}{1 - \chi_j'\left(p_{jn}(q)\right)},$$

and the second derivative is:

$$p_{jn}''(q) = \frac{\mu_n\ \chi_j''\left(p_{jn}(q)\right)\ p_{jn}'(q)}{\left(1 - \chi_j'\left(p_{jn}(q)\right)\right)^2} = \frac{\mu_n^2\ \chi_j''\left(p_{jn}(q)\right)}{\left(1 - \chi_j'\left(p_{jn}(q)\right)\right)^3}.$$

Given that $0 < \chi_j'(p) < 1, \forall p \in (0, \bar{p}_{jn})$ and $\chi_j''(p) > 0, \forall p > 0$ we conclude that $p_{jn}'(q) > 0$ and $p_{jn}''(q) > 0, \forall q \in (0, \bar{q}_{jn})$. ■

For a variety j, define function $h_j(q)$ as the difference between the expenditure of purchasing quality q in the informal sector and in the formal sector:

$$h_j(q) = p_{jn}(q)(1+i) - \min\{\mu_f(1+\tau)q(1+i), \mu_f(1+\tau)q + \gamma(j)\}$$

Optimal household behavior entails buying formal iff $h_j(q) \geq 0$. h_j is continuous.

Theorem 8 *If $p_{jn}'(q) > 0$ and $p_{jn}''(q) > 0$ $\forall q \in (0, \bar{q}_{jn})$, then $h_j''(q) > 0, \forall q \in (0, \bar{q}_{jn})$.*

Proof. The first derivative of h_j is:

$$h_j'(q) = \begin{cases} p_{jn}'(q)(1+i) - \mu_f(1+\tau)(1+i), & \text{if } \mu_f(1+\tau)q(1+i) < \mu_f(1+\tau)q + \gamma(j) \\ p_{jn}'(q)(1+i) - \mu_f(1+\tau), & \text{otherwise} \end{cases}$$

and the second derivative is:

$$h_j''(q) = p_{jn}''(q)(1+i).$$

Hence, $p_{jn}''(q) > 0$ implies $h_j''(q) > 0$. ■

Theorem 9 *If $h_j(0) = 0$ and $h_j''(q) > 0, \forall q \in (0, \bar{q}_{jn})$, then the threshold rule in Proposition 3 holds.*

Proof. There are three possible cases to consider:

1. If $h_j'(0) \geq 0$, then $h_j'(q) > 0, \forall q \in (0, \bar{q}_{jn})$ because $h_j''(q) > 0, \forall q \in (0, \bar{q}_{jn})$. Then, $h_j(q) > 0, \forall q \in (0, \bar{q}_{jn})$ and the household should buy from the formal sector regardless of quality. This corresponds to a case where the threshold $\hat{q}_j^n$ equals 0.
2. If $h_j'(0) < 0$ and $h_j(\bar{q}_{jn}) > 0$, the monotonicity of the first derivative, implied by the second derivative, means that there is a unique quality $\hat{q}_j^n > 0$ such that $h_j(\hat{q}_j^n) = 0$. Moreover, for $q < \hat{q}_j^n$, $h_j(q) < 0$ and for $q > \hat{q}_j^n$, $h_j(q) > 0$, proving the Proposition.
3. If $h_j'(0) < 0$ and $h_j(\bar{q}_{jn}) < 0$, there is no $q \in (0, \bar{q}_{jn})$ such that $h_j(q) = 0$, and $h_j(q) < 0, \forall q \in (0, \bar{q}_{jn})$. This correspond to a case where the threshold is the maximum quality available for sale in the informal sector, $\hat{q}_j^n = \bar{q}_{jn}$.

■

$h_j(0) = 0$ is trivially true, so the two conditions of the theorem above are satisfied as an implication from the assumptions made about $\chi_j(p)$.

We now prove that $\hat{q}_j^n$ is decreasing in j.

Theorem 10 *If $\chi_k(p) > \chi_j(p), \forall k > j$, then $\bar{q}_{jn} > \bar{q}_{kn}$.*

Proof. Our definition of $\bar{q}_{jn}$ can also be written as:

$$\bar{q}_{jn} = \max_p \frac{p - \chi_j(p)}{\mu_n},$$

and $\bar{p}_{jn} = \arg\max_p \, p - \chi_j(p)$. Then, we have:

$$\bar{q}_{kn} = \frac{\bar{p}_{kn} - \chi_k(\bar{p}_{kn})}{\mu_n} < \frac{\bar{p}_{kn} - \chi_j(\bar{p}_{kn})}{\mu_n},$$

where the inequality follows from $\chi_k(\bar{p}_{kn}) > \chi_j(\bar{p}_{kn})$.

Moreover, we have:

$$\bar{q}_{jn} = \max_p \frac{p - \chi_j(p)}{\mu_n} > \frac{\bar{p}_{kn} - \chi_j(\bar{p}_{kn})}{\mu_n}.$$

We conclude that $\bar{q}_{jn} > \bar{q}_{kn}$.

■

Theorem 11 *If $\chi_k(p) > \chi_j(p), \forall k > j$, then $p_{kn}(q) > p_{jn}(q), \forall q \in (0, \bar{q}_{kn})$.*

Proof. For any two varieties j, k, the following condition holds for good of quality q:

$$p_{jn}(q) - \chi_j(p_{jn}(q)) = p_{kn}(q) - \chi_k(p_{kn}(q)) \Rightarrow p_{jn}(q) - p_{kn}(q) = \chi_j(p_{jn}(q)) - \chi_k(p_{kn}(q))$$

Rewrite the right-hand side as:

$$\chi_j(p_{jn}(q)) - \chi_j(p_{kn}(q)) + (\chi_j(p_{kn}(q)) - \chi_k(p_{kn}(q)))$$

We proceed by contradiction. Suppose that $p_{jn}(q) \geq p_{kn}(q)$. Then, by the mean value theorem (relying on continuity and differentiability of χ_j) $\exists p^* \in [p_{kn}(q), p_{jn}(q)]$ such that:

$$\chi_j(p_{jn}(q)) - \chi_j(p_{kn}(q)) = \chi_j'(p^*)(p_{jn}(q) - p_{kn}(q))$$

Note also that $p^* < p_{jn}(q) \Rightarrow p^* < \bar{p}_{jn}$. Plugging the expression above into our previous condition:

$$p_{jn}(q) - p_{kn}(q) = \chi_j'(p^*)(p_{jn}(q) - p_{kn}(q)) + (\chi_j(p_{kn}(q)) - \chi_k(p_{kn}(q))),$$

and we find:

$$p_{jn}(q) - p_{kn}(q) = \frac{(\chi_j(p_{kn}(q)) - \chi_k(p_{kn}(q)))}{1 - \chi_j'(p^*)}$$

But because $p^* < \bar{p}_{jn}$, we have $\chi_j'(p^*) < 1$, so the numerator of the right-hand side of our equation is negative, while the denominator is positive, making the fraction negative. We get a contradiction, as we assumed that the left-hand side of the expression was positive. Hence, the contradiction proves that $p_{jn}(q) \geq p_{kn}(q)$ cannot be true, and its converse must hold. ■

Theorem 12 *If $\gamma(k) < \gamma(j), \forall k > j$ and $p_{kn}(q) > p_{jn}(q), \forall q \in (0, \bar{q}_{kn})$, then $h_k(q) > h_j(q), \forall q \in (0, \bar{q}_{kn})$.*

Proof.

$$h_k(q) - h_j(q) \geq [p_{kn}(q) - p_{jn}(q)](1 + i),$$

where the inequality follows from $\gamma(k) < \gamma(j)$. Then $p_{kn}(q) > p_{jn}(q)$ implies $h_k(q) > h_j(q)$.

■

Theorem 13 *If $h_k(q) > h_j(q), \forall q \in (0, \bar{q}_{kn})$, then $\hat{q}_j^n \geq \hat{q}_k^n$.*

Proof. Let's consider the different cases:

1. If $\hat{q}_j^n \in (0, \bar{q}_{jn})$, then $h_j(\bar{q}_j^n) = 0$. If $\hat{q}_j^n \leq \bar{q}_{kn}$, $h_k(\bar{q}_j^n) > 0$ and it must be the case that $\bar{q}_k^n < \bar{q}_j^n$. Alternatively, if $\hat{q}_j^n > \bar{q}_{kn}$, then because $\bar{q}_k^n \in [0, \bar{q}_{kn}]$, it follows that $\bar{q}_k^n < \bar{q}_j^n$.

2. If $\hat{q}_j^n = 0$, then $h_j(q) > 0, \forall q \in (0, \bar{q}_{jn}]$ and it follows $h_k(q) > 0, \forall q \in (0, \bar{q}_{kn}]$, implying $\hat{q}_k^n = \hat{q}_n^j = 0$.
3. If $\hat{q}_j^n = \bar{q}_{jn}$, then $h_j(q) < 0, \forall q \in (0, \bar{q}_{jn}]$. Given that $\bar{q}_{kn} < \bar{q}_{jn}$, $\hat{q}_k^n \in [0, \bar{q}_{kn}]$ implies that $\hat{q}_k^n < \hat{q}_j^n$.

■

Statement 2

The difference between the expenditure of purchasing formally variety j with quality q with cash and credit is given by:

$$\mu_f(1+\tau)q(1+i) - [\mu_f(1+\tau)q + \gamma(j)] = \mu_f(1+\tau)qi - \gamma(j)$$

Let $\hat{q}_j^m = \max\left\{\frac{\gamma(j)}{\mu_f(1+\tau)i}, \hat{q}_j^n\right\}$. The threshold rule follows. Given that both $\gamma(j)$ and $\hat{q}_j^n$ are decreasing in j, $\hat{q}_j^m$ is also decreasing in j.

Proposition 4

Define:

$$r^n(j,q) = q - \hat{q}_j^n.$$

This function is such that $r^n(j,q) \geq 0$ implies that a household purchasing q units would want to buy variety j formally. It would want to buy it informally otherwise. Because $\hat{q}_j^n \in [0, \bar{q}_{jn}]$ is decreasing in j, r^n is increasing in j. Trivially, the threshold rule follows, where one applies the intermediate value theorem for the case in which $\bar{j}^n(q)$ is interior, but it may also equal 1 if $r^n(j,q) < 0, \forall j \in [0,1]$ and the household buys all varieties informally, or it may equal 0 if $r^n(j,q) > 0, \forall j \in [0,1]$ and the household buys all varieties formally.

The same logic applies for the case of the method of payment threshold.

Goods market clearing

Define the measure of demand for formal goods of quality $q \in A$ for a Borel subset of $\mathbb{R}^+$, m^{D_f}, as equal to the left-hand side of the market clearing condition when $J = [0,1]$:

$$m^{D_f}(A) = \int_0^1 \int_0^1 \mathbb{I}\{q_{hjft} \in A\}\, dj\, dh + \int_0^1 \int_A g_{jt}(q)\, dq\, dj$$

First, note that a household h can instead be indexed by its state variables (x, ϵ), and the integral over h can be replaced by a integral over the appropriate space of state variables using the measure of households. Secondly, q_{hjf} is simply given by:

$$q_{hjf} = \begin{cases} q(x,\epsilon), & j \geq \hat{j}^n(q(x,\epsilon)) \\ 0, & otherwise \end{cases}$$

Hence, we can write our measure of demand as:

$$m^{D_f}(A) = \int \mathbb{I}\{q(x,\epsilon) \in A\}\Big[1 - \hat{\jmath}^n(q(x,\epsilon))\Big]\, d\Phi(x,\epsilon) + \int_0^1 \int_A g_j(q)\, dq\, dj$$

Similarly, define the measure of supply of formal goods of quality $q \in A$ for a Borel subset of $\mathbb{R}^+$, m^{S_f}, as equal to the right-hand side of the market clearing condition when $J = [0,1]$:

$$m^{S_f}(A) = \int_0^{\nu^F} \left(\mathbb{I}\{g_{kt} = f\} \int_A s_{kt}(q)\, dq\right) dk$$

A firm k can be indexed by its productivity z, and the integral over firm index can be replaced with an integral over the probability distribution of productivity. Moreover, a firm will formalize iff $z \geq \hat{z}^n$. Let $s(z,q)$ be the supply of quality q by a firm with productivity z, replacing $s_k(q)$. Thus,

$$m^{S_f}(A) = \nu^F \int_{\hat{z}^n}^{\bar{z}} \int_A s(z,q)\, dq \;\; d\Psi(z)$$

Now, we have that the two measures are equal for any set A, and we can write:

$$\int_0^{+\infty} q\, dm^{D_f}(q) = \int_0^{+\infty} q\, dm^{S_f}(q)$$

This yields:

$$\int \Big[1 - \hat{\jmath}^n(q(x,\epsilon))\Big] q(x,\epsilon)\, d\Phi(x,\epsilon) + \int_0^1 \int g_j(q) q\, dq\, dj = \int_{\hat{z}^n}^{\bar{z}} \int s(z,q) q\, dq \;\; d\Psi(z)$$

Using $p_{jf}(q) = \mu_f(1+\tau)q$, we can manipulate our expression for government expenditure:

$$G_t = \int \int \frac{p_{jf}(q)}{1+\tau} g_j(q)\, dq\, dj \implies \frac{G_t}{\mu_f} = \int \int g_j(q) q\, dq\, dj$$

Using the capacity constraint, we know that for a formal firm:

$$\int s(z,q) q\, dq = F(z, \ell_f(z))$$

Hence, we get:

$$\int \Big[1 - \hat{\jmath}^n(q(x,\epsilon))\Big] q(x,\epsilon)\, d\Phi(x,\epsilon) + \frac{G_t}{\mu_f} = \nu^F \int_{\hat{z}^n}^{\bar{z}} F(z, \ell_f(z)) \;\; d\Psi(z)$$

This shows that the market clearing condition for formal goods of different qualities and varieties implies our aggregate sector level condition. The same holds for the case of informal goods.

One also needs to show that the reverse is true, that is, that given the aggregate condition and price schedules one can construct an equilibrium in which market clearing conditions for goods of different qualities and varieties hold. When the equilibrium price schedules holds, the measure of supply of a set of qualities A of varieties J is undetermined because firms get the same profits by producing any variety or quality composition. Hence, we can construct the case in which the measure of supply exactly matches the measure of demand for all possible sets. Having done this, our aggregate conditions are satisfied by the proof above. Hence, it follows that when the aggregate conditions are satisfied and equilibrium price schedules hold, there is an equilibrium.

C Model solution and additional results

Computation of the household problem

1. Define a grid for quality: $\mathcal{G}^q = [q_1, q_2, ..., q_{N_q}]$ and a grid for varieties $\mathcal{G}^j = [j_1, j_2, ..., j_{N_j}]$.
2. For each variety $j_l \in \mathcal{G}^j$ compute the quality thresholds $\hat{q}^n_{j_l}$ and $\hat{q}^m_{j_l}$.
   (a) Solve the equation $\mu_f(1+\tau)q(1+i) - \gamma(j_k) = 0$ to find $\tilde{q}^m_{j_l}$.
   (b) Solve the equation $p_{j_l n}(q)(1+i) - \mu_f(1+\tau)q(1+i) = 0$, and denote the solution as $\tilde{q}^n_{j_l}$.
   (c) If $\tilde{q}^n_{j_l} \leq \tilde{q}^m_{j_l}$, then $\hat{q}^n_{j_l} = \tilde{q}^n_{j_l}$ and $\hat{q}^m_{j_l} = \tilde{q}^m_{j_l}$. If $\tilde{q}^n_{j_l} > \tilde{q}^m_{j_l}$, $\hat{q}^n_{j_l}$ is the solution to the equation $p_{j_l n}(q)(1+i) - \gamma(j_l) = 0$ and $\hat{q}^m_{j_l} = \hat{q}^n_{j_l}$.
3. For each quality $q_k \in \mathcal{G}^q$, compute the threshold varieties, $\hat{j}^n(q_k)$ and $\hat{j}^m(q_k)$, and their respective indices, $\hat{l}^n(q_k)$ and $\hat{l}^m(q_k)$
   (a) Find the largest j_l for which $q_k - \hat{q}^n_{j_l} < 0$, this yields $\hat{j}^n(q_k)$.
   (b) Find the largest j_l for which $q_k - \hat{q}^m_{j_l} < 0$, this yields $\hat{j}^m(q_k)$.
4. For each quality $q_k \in \mathcal{G}^q$, compute the total expenditure:

$$E(q_k) = \sum_{l=1}^{\hat{l}^n(q_k)} (1+i)p_{j_l n}(q_k) + \sum_{l=\hat{l}^n(q_k)+1}^{\hat{l}^m(q_k)} \mu_f(1+\tau)(1+i)q_k + \sum_{l=\hat{l}^m(q_k)+1}^{N_j} [\mu_f(1+\tau)q_k + \gamma(j_l)]$$

5. Interpolate the expenditure function computed in the grid $\mathcal{G}^q$.
6. Solve the dynamic programming problem of the household:

$$\begin{aligned} V(x,\epsilon) = \max_{q\in[0,\bar{q}],l\geq 0} u(q,l) + \beta\mathbb{E}V(x',\epsilon'), \\ st: x' = \left(\frac{1+i}{1+\pi}\right) x + \epsilon l - E(q), \\ E(\bar{q}) = \left(\frac{1+i}{1+\pi}\right) x. \end{aligned}$$

Stationary Equilibrium

The Government picks three variables out of (i, τ, G, π, D)[19] and the two other will be endogenous (i and π cannot both be chosen because of the exogenous real rate condition). For calibration, we set D,τ, and π using data and let i and G be equilibrium outcomes. In the main quantitative exercise, we set D, G, and π and let i and τ be equilibrium outcomes. The general algorithm for computation is as follows:

1. Guess a value for each of the variables in the set $\{\tau, \mu_f, \mu_n\}$ which are endogenous.
2. Compute i to match the exogenous real rate.
3. Compute the productivity threshold $\hat{z}^n$ for the formalization decision.
4. Compute the labor demand of firms $(\ell_f(z), \ell_n(z))$.
5. Compute price schedules $(p_{jf}(q), p_{jn}(q))_j$.
6. Solve the household problem, getting policy functions.
7. Find the stationary distribution of households.
8. Use the money market clearing condition to get aggregate money M.
9. Compute the aggregate bond supply using D and M and then compute tax revenues and use the law of motion of government assets to compute implied G.
10. Use the formal sector goods market condition to compute the mass of firms ν^F.
11. Check the errors in the informal sector goods market clearing condition, free entry condition, and government expenditure if G is exogenous. If error is small, conclude. If not, update guess in step 1 and repeat.

Data and moments

We use Peru's household survey (ENAHO 2016-2020, panel version) to compute moments on households' earnings processes and firm size dispersion. For measures of the size of the informal economy we use the Informal Economy Database (Elgin et al., 2021). Most of the remaining empirical targets are computed as averages between years 1990 and 2016 of macro data series retrieved from either the World Bank or the Peruvian Central Bank.

- *Household earnings process*: We compute the dispersion and the (yearly) autocorrelation of log earnings. To compute the autocorrelation parameter of the quarterly $AR(1)$ process, we raise the year autocorrelation by 1/4.
- *Firm size dispersion*: we approximate the firm size distribution using the firm size reported by the workers surveyed in the ENAHO. To estimate the number of firms of size

[19] π and g_N are interchangeable in the Stationary Equilibrium. While strictly speaking what the Government chooses is the growth rate of nominal liabilities, this pins down the inflation rate, so we can treat it as a direct choice over π. $D = B/(1+i) + M$ is the total amount of government liabilities which we think of as representing government debt.

S we divide the number of workers that report working in firms of size S by S. This corrects for the higher probability of sampling a worker of a larger firm. We compare our firm size distribution with the one reported by the Ministry of Production for the share of micro, small, medium and large enterprises and find that it matches this well. The standard deviation of firm size (in logs) equals 0.663.

- *Semi-elasticity of money demand*: we replicate the methodology of Dotsey and Ireland (1996) using the IMF's IFS dataset for years 1995-2016.
- *CPI inflation and interest rate*: these two series are retrieved the World Bank. The interest rate corresponds to the real deposit rate, which is the standard return paid by commercial banks for time and savings deposits.
- *Government debt and monetary aggregates*: these series are retrieved from the Peruvian Central Bank. For money we use the M0 series named 'circulante'.

Additional results: policy change in economies with different level of inequality

An implication of the framework described in Section 2 is that inequality amplifies the distributional concerns associated with the non-homothetic demand for informal consumption. Higher consumption inequality steepens informality Engel curves, which in turn magnifies differences in effective tax rates and relative tax incidence across expenditure groups.

Table 6
The role of inequality

		Benchmark	Low inequality	High inequality
Welfare change (%)	P25	-0.21	0.06	-0.44
$\pi = 4\% \rightarrow \pi = 10\%$	P50	-0.04	0.20	-0.26
	P75	0.11	0.32	-0.09
Informal discount (%)	P25	10.1	9.5	10.9
	P50	8.3	7.9	8.6
	P75	7.1	7.0	7.2
Preferred inflation (%)	P25	10.1	9.5	10.9
	P50	8.3	7.9	8.6
	P75	7.1	7.0	7.2

Table 6 quantifies this by contrasting aggregate statistics of the benchmark model with low- and high-inequality counterfactuals. We vary consumption inequality through the volatility of household productivity: the low (high) inequality case features a 25% lower (higher) dispersion of log earnings. Moving from the low to the high inequality counterfactual increases the interquartile range of the informal discount from 2.5% to 3.7%. Analogously, the interquartile range of the welfare gains of moving from 4% to 10% inflation rises from 0.26 to 0.35.

These results illustrate that consumption inequality has first-order implications for the equity consequences of indirect taxation. This is particularly relevant in developing countries, where larger informal sectors and heavier reliance on indirect taxes coincide with higher levels of inequality (Chong and Gradstein, 2007).