

A Model of Firm Formality and Tax Evasion

Kunal Sangani*

Undergraduate Honors Thesis
Stanford University

May 5, 2016

Abstract

Rampant informality in developing economies proves to be a major obstacle to public revenue collection, limiting investment in crucial public infrastructure. The model presented in this paper sheds light on inter-firm mechanisms that motivate firms to either transact formally or transact informally and evade taxes. The formulation allows us to predict comparative statics on a variety of parameters, including the effective tax rate, the likelihood of audit, the penalties levered against firms caught transacting informally, and the distribution of firm sizes in the economy. Empirical support for each of these predictions is discussed. In particular, the model provides theoretical underpinnings for predicting the effect of multinational entry and trade liberalization on the size of the informal sector—two questions which have hitherto been largely neglected in the literature.

1 Introduction

As Nicholas Kaldor (1963) put forward in his seminal work on public finance, “the importance of public revenue from the point of view of accelerated economic development could hardly be exaggerated.” A state’s development is limited and enabled by that state’s power to tax. In particular, states with a strong tax base are able to make investments in public infrastructure that improve standard of living and set the stage for economic growth.

*Email address: ksangani@stanford.edu. I’d like to thank Marcelo Clerici-Arias for his constant support as organizer of the honors program and my major advisor, Fuhito Kojima, for his advice throughout my time at Stanford. I’d also like to thank my friends and family for helpful comments on drafts. I am particularly indebted to my thesis advisors, Matthew Jackson and Michael Best, for their time, comments, advice, and inspiration, without which I would not have been able to write this paper.

It is crucial, then, to understand the factors that limit collection of tax revenues. Informality—defined broadly as unregistered economic activity that would otherwise be taxable—is rampant in many countries, especially in those where revenue for public infrastructure may be most necessary. La Porta and Shleifer (2014) report the number of registered firms per thousand citizens increases from 3.2 in the bottom quintile of countries by per capita income to 41.8 in the top quintile of countries; the level of tax evasion across these quintiles decreases from 29% to 8%. The question at the heart of this paper is: What factors influence a firm’s decision to either transact formally or to transact informally and evade taxes? Shedding light on firms’ internal decision-making around formality promises predictions for how tax policy and exogenous factors change the extent of informality in an economy in aggregate.

In addition to standard parameters—the effective tax rate, the likelihood of audit, and the penalty rate for tax evasion—the model presented here emphasizes opportunities for trade with other firms in the economy as a key motivation for firm decisions regarding formality. Given the risk to formal firms of transacting with firms in the informal economy, firms choose to transact formally or informally not only due to the burdens of tax or the ease of slipping under government radar, but also because of the opportunities available to them in the pool of firms they enter. Stressing the importance of these opportunities in firms’ decision-making has the additional advantage of allowing comparative statics arguments on the impact, for instance, of multinational entry and trade liberalization on the size of the informal sector.

There is a relative dearth of both theoretical and empirical literature on the impact of multinationals on informality and tax evasion in the economies they enter. Generally, there appears to be a pessimism around multinational entry into developing countries following the exposé of bribing practices by Walmart’s extension to Mexico in 2012.¹ This cynicism is furthered by arguments that multinationals increase the extent of informality in economies they enter: By competing out firms in the formal economy of the host country, multinationals displace workers from the formal sector and feed the growth of informal firms that are relatively unproductive and too small to achieve economies of scale.² On the other hand, some—proponents of the Foreign Corrupt Practices Act (FCPA), for instance—argue that multinationals bring about norms that discourage bribery, tax evasion, and other forms of corruption. This is precisely the converse of the Walmart case: steep legal penalties and media scrutiny in the multinational’s home country discourage misbehavior, and norms in the host country evolve to meet the multinational’s standard. In either case, the effect of multinationals on informality has not been adequately substantiated by theoretical models or empirical studies.

More literature has been dedicated to trade liberalization, though there too Fugazza and Fiess (2010) note that “the relationship between trade liberalization and informality has not

¹See the New York Times coverage of the Walmart scandal at <http://www.nytimes.com/2012/12/18/business/walmart-bribes-teotihuacan.html>.

²Whether cynicism around multinational entry as a whole is substantiated is not the focus of this paper. Rather, we evaluate the narrow claim that multinational entry increases the extent of informality. In Section 5, we discuss the need to understand broader welfare impacts that are outside the scope of the model presented here.

received the attention, whether theoretical or empirical, that it may deserve.” We define trade liberalization broadly as the removal of both tariff and non-tariff obstacles on trade between nations. Such a removal increases not only the size of the effective market for a firm but also the pool of firms with which a firm is interacting or transacting with. For firms in developing markets, trade liberalization allows contact with a significantly greater number of large firms based in developed markets. The literature has posited both positive and negative effects of trade liberalization on the extent of informality: Neither hypothesis has been clearly validated.

The model presented in this paper, then, not only illustrates a novel mechanism by which firms choose between formality and informality, but also gives theoretical underpinnings to unanswered questions about the size of the informal economy. We start with a review of related literature, identifying the unique mechanism our model captures among models of tax compliance in Section 2. In Section 3, we present the model in both a specified and general form, and proceed to discuss the model’s implications and empirical support in Section 4. Section 5 concludes the paper with a discussion of other factors, outside the scope of this model, that should be given attention in future work stemming from this paper.

2 Literature Review

This paper draws from a detailed literature on informality and tax compliance. The literature grows largely from a seminal model of tax evasion developed by Allingham and Sandmo (1972). In the A-S model, rational agents choose to pay taxes only when the threat of audit outweighs the benefit of evading taxes. When both the likelihood of government audit and the associated penalties are low—which they tend to be in practice—the standard model predicts low compliance. However, Adreoni, Erard and Feinstein (1998) note that “a small amount of enforcement has gone a fairly long way”: IRS data suggests that only about 17 percent of taxes go uncollected in the U.S., where the rate of returns being audited is 1 percent. The authors comment that this is one of the fundamental puzzles in the field: “The most significant discrepancy that has been documented [is that] the theoretical model greatly over-predicts noncompliance.”

Recent studies have argued that high compliance may be in part due to behavioral and psychological aspects linked to tax-paying, such as patriotism, social norms, guilt, and tax morale. Yet, Slemrod, Blumenthal and Christian (2001) find in a field experiment conducted with the Minnesota Department of Revenue that normative appeals sent to a group of taxpayers had no significant effect on compliance, while letters indicating a threat of audit in the coming year significantly increased compliance among taxpayers in most income classes. Other studies too have reinforced the importance of the standard parameters in the A-S model on tax evasion: Beron, Tauchen and Witte (1992) find that the odds of an audit increase reported tax liability for some, though not all, groups in IRS data aggregated by the TCMP, and Feinstein (1991) finds that the marginal tax rate is negatively associated with evasion in a cross-sectional analysis of TCMP data.³

³A more thorough review and critical analysis of these and related studies are available in the chapter on

Importantly, these empirical studies find that noncompliance is most rampant in areas where income is self-reported. Slemrod, Blumenthal and Christian (2001), for instance, find that letters indicating a threat of audit have little or no effect on the level of tax-paying for those with third-party reported incomes. The increase in compliance that followed threats of audit was due solely to increased tax-paying by those with self-reported incomes, a conclusion corroborated by Kleven *et al* (2011) in a tax enforcement field experiment conducted in Denmark. This empirical fact has led some researchers to attribute the unexpectedly low levels of evasion observed in most developed countries to the existence of third-party reporting. Sandmo (2005) and Slemrod (2007) note the likelihood of audit is in fact a poor proxy for the likelihood of tax fraud detection when third-party reporting of information exists, since mismatched reports easily reveal cheating. Information, when reported by multiple parties, appears to play a critical role in maintaining a high rate of compliance.

Of course, the high tax compliance observed in most developed countries is not ubiquitous in all economies. Evasion reaches extremely high levels in some developing countries: Krugman *et al* (1992) report that as few as 50% of income taxes were paid in Philippines in 1985. Tax compliance is not static within countries either: Madisson (2001) reports that the tax capacity of France, Germany, the Netherlands, and the UK increased from 12% of GDP to 46% from 1910 to 2000. A central question, then, is what specific mechanisms make third-party reporting an effective measure for high tax compliance, and how these mechanisms depend on the characteristics of the private sector and government agencies in the context of a country's development.

A few recent papers have shed some light on this question by offering models of how third-party reporting of information encourages tax compliance. These models can largely be separated into two classes of mechanisms: those which emphasize the role of the within-firm information network and those that emphasize the role of the across-firm information network. The first class is best represented by Kleven, Kreiner and Saez (2015), who present a model of a firm with employees who may accidentally or deliberately reveal collusive tax cheating to the government. Embedded within an endogenous growth model, the firm's use of business records increases as it hires more employees. Kleven, Kreiner and Saez conclude that due to the increased ease of whistle-blowing as the firm grows, the government is able to enforce taxes despite a low threat of audit.

This model illustrates the effect of within-firm business records and increased employee counts on compliance. It does not necessarily explain, however, the high level of compliance observed for small- and mid-sized firms in developed nations. If compliance were merely a byproduct of the risks of evasion when employee counts are high and business records are extensive, this does not account for the fact that firms of a certain size may be almost ubiquitously formal in one country and informal in another.

Explanations that emphasize the role of networks across firms in increasing compliance are, therefore, a useful complement. Kopczuk and Slemrod (2006) comment on the importance of inter-firm interaction in their comparison of value-added tax (VAT) and retail sales tax (RST) schemes on the level of tax collected from corporate income, but do not ascribe

Tax Avoidance, Evasion, and Administration by Slemrod and Yitzhaki (2002).

the inter-firm collusion costs they include to a specific mechanism. Gordon and Li (2009) argue that links to the financial sector discourage informality due to the bookkeeping required in the financial services. In a field experiment conducted in conjunction with the Chilean Tax Authority, Pomeranz (2013) finds evidence that the inter-firm paper trail generated by a VAT scheme indeed has a preventative deterrence effect on tax evasion, a conclusion supported by de Paula and Scheinkman (2010), who find that the tendency of one firm to evade taxes is correlated to the informality of firms both up- and downstream in production. These papers identify aspects of the inter-firm network that may increase compliance, but are shy of identifying the critical channels by which reporting of information dissuades evasion and how those channels are impacted by properties of the private sector.

Acemoglu and Jackson (2014) present a model of social norms and law enforcement in which agents are randomly matched and benefit from exhibiting behaviors similar to their partners. Law-abiders paired with law-breaking agents may whistle-blow, thus dissuading agents inclined to break the law *ex ante* when social norms dictate the observation of the law. In a tax context, this corresponds to a mechanism of social norm adherence and whistle-blowing between transacting firms that encourages tax compliance.

The model in this paper borrows from the technology used in the paper by Acemoglu and Jackson, with some fundamental differences. For one, the primary mechanism that encourages compliance in this paper’s model is not whistle-blowing, but instead the opportunities available to a firm given their decision to be either formal or informal. Firms enter either the formal or informal sector of the economy after considering the transaction opportunities available in both. This model has the advantage of characterizing multiple equilibria and provides opportunities to understand the result of an exogenous shock—such as the entrance of a large multinational into the private sector of a developing economy—on norms and compliance.

Note that the model given here follows in the vein of de Paula and Scheinkman (2010) by defining informality as complete tax evasion. Since the broader motivation of this paper is to understand how parameters shaping informality contribute to the collection of public revenues, this is the most obvious definition. However, the conclusions of this model are applicable to broader definitions of informality as unregistered or underground activity. For instance, the parameter describing the tax rate in this model can be seen as a proxy for the costs of registering formally and complying with regulation. We use this understanding when discussing the empirical support and applications of our model to draw from literature that may include broader definitions of informality.

In the next section, we present this model and the comparative statics it predicts; the discussion, which follows, matches these predictions to empirical findings available in the literature and considers broader implications.

3 Model

Suppose the economy consists of a set of firms, N . Each firm $i = 1, \dots, n$ has size $s_i \in (0, 1]$, such that the distribution of firm sizes in the economy follows the density function $f(s)$. A

firm makes a binary decision to either transact formally ($a_i = 1$) or informally ($a_i = 0$).⁴ A function $m : N \rightarrow N$ then matches each firm with another firm in the market such that

- (a) $i = m(j)$ iff $j = m(i)$,
- (b) $i \neq m(i)$

We assume that n is even, so that each firm has a match. If both firms are formal or both firms are informal, then the pair of firms transacts, so that firm i enjoys a profit given by $\pi(s_i, s_{m(i)})$. If there is a mismatch in formality, however—firm i is formal while firm $m(i)$ is informal, or vice versa—then the firms do not transact with each other, and the resulting payoff for both firms is zero.⁵

For firms that do transact, firms that have chosen to be formal pay a tax rate τ on this profit, such that the firm i 's payoff is given by:

$$u_i(a_i = 1; s_i) = \mathbb{1}(a_{m(i)} = 1) \pi(s_i, s_{m(i)})(1 - \tau)$$

Firms that choose to be informal are subject to being caught by government auditors. The probability that an informal transaction is caught depends on the sizes of the two firms transacting, and is given by $\mu(s_i, s_{m(i)})$. If caught, a firm pays penalty rate T on its profit. This gives us the payout for firm i :

$$u_i(a_i = 0; s_i) = \mathbb{1}(a_{m(i)} = 0) \pi(s_i, s_{m(i)})(1 - T\mu(s_i, s_{m(i)}))$$

Combining the above gives us the complete payoff function for firm i :

$$u_i(a_i; s_i) = \begin{cases} \pi(s_i, s_{m(i)})[1 - a_i\tau - (1 - a_i)T\mu(s_i, s_{m(i)})] & \text{if } a_i = a_{m(i)} \\ 0 & \text{if } a_i \neq a_{m(i)} \end{cases}$$

The payoff function given in this model immediately yields two equilibria: one in which all firms are formal, and one in which all firms are informal. The all-formal and all-informal equilibria are stable as long as the tax rate and expected audit penalty are not prohibitively costly for any firm, i.e. $\tau < 1$ and $E_j T\mu(s_i, s_j) < 1$ for all s_i .

This model allows us to analyze how the extent of informality is affected by an array of parameters: the tax rate, τ ; the likelihood of audit described by μ and the resulting penalty rate, T ; and the distribution of firm sizes in the economy, f . Proofs for all following lemmas and propositions are included in the appendix.

⁴While it is true that tax evasion in some firms takes the form of mis- or underreporting taxes, we make a stricter assumption of complete formality or informality for the purpose of our model.

⁵The underlying assumption here is that formal and informal firms do not transact with one another. This is not an overly restrictive assumption: Böhme and Thiele (2014) find in a survey of informal West African entrepreneurs that both backward and forward linkages from from the informal sector to the formal sector are rare, and that the prevalence of these linkages is especially low when firms are unregistered. We intuit that formal firms who are completely truthful when reporting transaction records to the government are unwilling to transact with informal partners.

3.1 Equilibria for specified π, μ

To illustrate the model's predictions, we first analyze the equilibria possible when we specify functional forms for π and μ . In particular, we set

- (a) $\pi(s_i, s_{m(i)}) = s_i s_{m(i)}$, and
- (b) $\mu(s_i, s_{m(i)}) = \rho \max\{s_i, s_{m(i)}\}$

These functional forms assume that the profit from transaction between two firms increases with the sizes of both firms and that the probability a tax collection agency catches an informal transaction is proportional to the size of the larger firm involved.

Given these functional forms, we first analyze a simple economy in which firms all have one of two possible sizes, then proceed to analyze an economy with an arbitrary distribution of firm sizes.

3.1.1 Two firm sizes

Suppose our economy consists of firms of only two possible sizes, s_1 and s_2 with $s_1 > s_2$. There are m firms of size s_1 , and the remaining $N - m$ firms are of size s_2 . As shown above, two stable equilibria exist: the all-formal and all-informal equilibria. The following proposition tests the existence of equilibria in which some firms transact formally and others informally.

Proposition 3.1 *An equilibrium in which all large firms are formal and all small firms are informal is possible for a range of tax rates, τ , such that*

$$\frac{s_2[1 - \rho T s_1]}{s_1} \left(\frac{N - m}{m - 1} \right) \leq (1 - \tau) \leq \frac{s_2[1 - \rho T s_2]}{s_1} \left(\frac{N - m - 1}{m} \right)$$

There is no equilibrium in which all large firms are informal and all small firms are formal.

We find that there is an equilibrium in which small firms are informal and large firms are formal, but not vice versa. This result is enabled by two forces: First, our form for μ implies that large firms are at greater risk of penalty when they attempt to transact informally. Intuitively, we expect larger firms to have operations that may be more easily spotted by tax collection officials, leaving them exposed to audit. Drawing from the model by Kleven, Kreiner and Saez (2015), we might also expect that firms need to become more regimented and keep better business records as the number and size of operations grow; the resulting paper trail makes large firms easier to detect. Second, there is no countervailing upside in profit from informality that ballasts against the increased risk of penalty for large firms attempting to avoid detection. Since the underlying profit opportunity for large firms transacting informally or formally is not substantially different, and the expected penalty for transacting informally increases with firm size, we arrive at a result where large firms favor formality and small firms informality.

Empirical evidence shows that formality is indeed positively correlated with firm size, and the simple case above provides an intuition for this stylized fact. This finding is discussed in more detail in Section 4. We now turn to a continuous distribution of firm sizes to demonstrate how this intuition holds in the continuous case.

3.1.2 Arbitrary distribution of firm sizes

Suppose the distribution of firms in the economy is instead given by the probability density function $f(s)$.⁶ Extrapolating from the above case, we expect larger firms to tend toward formality and smaller firms to be prone to informality.

Lemma 3.2 *If a firm of size s weakly prefers transacting formally to transacting informally, then a firm of size $s' > s$ strictly prefers transacting formally to transacting informally.*

This lemma shows that the preference for formality increases with firm size. In particular, it suggests the presence of a cutoff equilibrium: all firms smaller than some cutoff size choose to transact informally and all firms greater than that cutoff size choose to transact formally. Solving for this cutoff point, we find:

Proposition 3.3 *There exists an equilibrium with cutoff value c in which all firms of size $s_i < c$ do not pay taxes ($a_i = 0$) and all firms of size $s_i \geq c$ pay taxes ($a_i = 1$), where*

$$c = \frac{1}{\rho T} \left(1 - (1 - \tau) \frac{(1 - F(c))E[s_i | s_i > c]}{F(c)E[s_i | s_i < c]} \right)$$

and where F is the cumulative distribution of firm sizes. For $c \in (0, 1)$, we must have τ such that

$$(1 - \rho T) \frac{F(c)E[s_i | s_i < c]}{(1 - F(c))E[s_i | s_i > c]} < (1 - \tau) < \frac{F(c)E[s_i | s_i < c]}{(1 - F(c))E[s_i | s_i > c]}$$

We have shown that the conclusion from our two firm size case extends to an arbitrary, continuous distribution of firm sizes: under certain conditions, we find an equilibrium in which small firms transact informally and large firms transact formally. Formal and informal sectors of the economy co-exist, and a firm's participation in either sector is dictated by the firm's size.

⁶Note that, with a continuous distribution of firm sizes, we run into measurability problems given our random matching function m . Feldman and Gilles (1985) and Judd (1985) provide a discussion of these issues for a continuum of agents with independent observations. In this model, though, independence of firm sizes is not needed. Applying the argument given by Barbera and Jackson (2016), we can formalize the firm sizes, s_i , drawn by firms by randomly assigning an agent to have the lowest signal drawn from $[0, 1]$, distribute signals in a nondecreasing way for agents with higher labels, and wrap around at 0. This process gives the correct distribution of firm sizes without measurability issues. Furthermore, independence of types is not needed since each firm cares only about the behavior of the population and not that of particular agents.

While the above case substantiates our intuition on the correlation between formality and firm size, comparative statics on how the cutoff solution changes with the model's parameters are not particularly tractable. We turn to the general case to assess how the above conclusions hold when we loosen our restrictions on the functional forms of π and μ and to present comparative statics for the tax rate, the penalty rate, the likelihood of audit, and the distribution of firm sizes.

3.2 Equilibria for general π, μ

We begin by imposing conditions on π and μ to prove a lemma analogous to lemma 3.2 above.

Lemma 3.4 *If a firm of size s weakly prefers transacting formally to transacting informally, then a firm of size $s' > s$ strictly prefers transacting formally to transacting informally under the following conditions:*

- (a) π can be written as $\pi(s_i, s_j) = \pi_1(s_i)\pi_2(s_j)$,
- (b) π_1 and π_2 are strictly positive, and
- (c) $\mu(s_i, s_j)$ is strictly increasing in its first argument.

Since we expect the profit from transacting with another firm to be positive, the assumption that π_1 and π_2 are strictly positive agrees with our model's formulation. The assumption that μ is strictly increasing in its first argument also appears to be reasonable, as we expect larger firms to have more difficulty in staying under the government radar when engaging in off-the-books transactions. The first condition is perhaps the most restrictive; it allows, still, for a wide variety of functional forms for π .⁷

Lemma 3.4 provides evidence for a cutoff solution. With this intuition, we can reformulate the game to remove firm types. Suppose that firm i chooses some cutoff size ex ante, c_i . Firm i then draws some size $s_i \in (0, 1]$ from the density function f and proceeds to transact formally if $s_i \geq c_i$ and transact informally if $s_i < c_i$. Modifying the game in this way gives us the payoff function for firm i :

$$u_i(c_i) = \begin{cases} \mathbf{1}(a_{m(i)} = 1) \pi(s_i, s_{m(i)})[1 - \tau] & \text{if } s_i \geq c_i \\ \mathbf{1}(a_{m(i)} = 0) \pi(s_i, s_{m(i)})[1 - T\mu(s_i, s_{m(i)})] & \text{if } s_i < c_i \end{cases}$$

In expectation, the firm's payoff is then:

$$E[u_i(c_i)] = \Pr(s_i \geq c_i) \Pr(a_{m(i)} = 1) E[\pi(s_i, s_{m(i)})[1 - \tau]] \\ + \Pr(s_i < c_i) \Pr(a_{m(i)} = 0) E[\pi(s_i, s_{m(i)})[1 - T\mu(s_i, s_{m(i)})]]$$

⁷The conditions given in lemma 3.4 are sufficient but not necessary. The third condition, for instance, is a stronger form of the requirement that $E_j \mu(s_i, s_j)$ is strictly increasing in s_i . At first glance, the functional form specified for μ in Section 3.1 does not satisfy the third condition, but it does satisfy this softer requirement.

If all other firms are playing some identical cutoff c , i.e. $c_j = c$, $\forall j \neq i$, we can rewrite firm i 's expected payoff as:

$$E[u_i(c_i)] = (1 - F(c_i))(1 - F(c))E_{s_i \geq c_i, s_j \geq c}[\pi(s_i, s_j)][1 - \tau] \\ + F(c_i)F(c)E_{s_i < c_i, s_j < c}[\pi(s_i, s_j)][1 - T\mu(s_i, s_j)]$$

where F is the cumulative distribution function corresponding to density function f .

Our formulation of the model posits that firms tend to gravitate toward the norms of formality in the wider pool of firms. This intuition suggests that the game may be supermodular, and indeed we confirm that it is as long as the tax rate and expected penalty are not prohibitively high for any firm:

Lemma 3.5 *The game is supermodular if $\pi(s_i, s_j) \geq 0$ and $\tau + T\mu(s_i, s_j) \leq 2$ for all s_i, s_j .*

Supermodularity indicates that each firm's best response cutoff is increasing in the cutoffs of other firms in the economy. When the conditions for supermodularity hold, we can apply comparative statics results from Milgrom and Roberts (1990) and Zandt and Vives (2007).

In general, when π and μ are unrestricted, our game may have multiple equilibria. The comparative statics results cited above apply only to our minimum, maximum, and stable intermediate equilibria; intermediate, unstable equilibria move in the opposite direction. Note that our extremal equilibria, in which all firms play $c = 0$ or $c = 1$, are stable and fixed. In simulations included in the following section, we illustrate how best response curves form in a representative case. We find that best response curves behave nicely, forming a unique, unstable equilibrium between the fixed all-formal and all-informal equilibria.

For this case, we formulate our results in terms of dynamic play as follows: Suppose some norm cutoff is drawn from a uniform distribution over $[0, 1]$. Firms begin by playing this cutoff, then play strategically and converge to either the all-informal or all-formal equilibrium. As the intermediate equilibrium moves to the right, the likelihood of converging to the all-formal equilibrium ($c_i = 0$ for all i) increases; as the intermediate equilibrium moves to the left, the likelihood of converging to the all-formal equilibrium decreases.⁸

In this dynamic play, we lose the conclusion from the above specified case in which informal and formal economies co-exist. In its place, though, we are able to make conclusions about how the likelihood of converging to the all-formal equilibrium—a proxy for the norm and extent of informality—changes with manipulations of our model's parameters. When the conditions for supermodularity hold and when we have a unique intermediate equilibrium, we predict the following comparative statics:

⁸Note that, alternatively, we can unfix our extremal equilibria from the all-formal and all-informal cases by including some types of firms that are "locked-in" to being informal or formal ex ante. This version of the game predicts stable cutoff equilibria that move in the same direction as the likelihood of converging to the all-formal equilibrium given in the propositions here. More detail on this alternative formulation is included in the appendix.

Proposition 3.6 *The likelihood of convergence at the all-formal equilibrium is decreasing in τ .*

Proposition 3.7 *If $\mu(s_i, s_j) \geq 0$ for all s_i, s_j , the likelihood of convergence at the all-formal equilibrium is increasing in T .*

Proposition 3.8 *If we have some probability of audit function $\mu'(s_i, s_j) = \rho\mu(s_i, s_j)$ where $\rho > 1$, and if $\mu(s_i, s_j) \geq 0$ for all s_i, s_j , then the likelihood of converging at the all-formal equilibrium under audit likelihood μ' is greater than the likelihood of converging at the all-formal equilibrium under μ .*

Proposition 3.9 *If there are two cumulative distribution functions of firm sizes F_1 and F_2 such that F_1 first order stochastically dominates F_2 , then the likelihood of converging to the all-formal equilibrium is greater under distribution F_1 than the likelihood of converging to the all-formal equilibrium under F_2 .*

These predictions—their empirical support as well as their real-world applications—are discussed in Section 4.

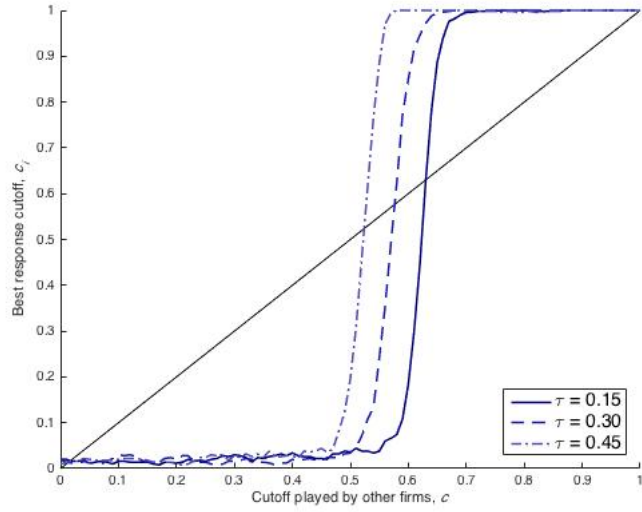
3.3 Simulations

To illustrate the predictions made by this model, we simulate best response cutoff curves for a firm i given cutoffs played by all other firms in the economy. Note that the parameters and the magnitudes of the predicted effects are not meant to be calibrated by empirical data. Rather, the simulations indicate the directional consequences of changes in the parameters included in the model.

For simulations, we use the profit and audit functional forms given in the specified case: $\pi(s_i, s_j) = s_i s_j$ and $\mu(s_i, s_j) = \rho \max\{s_i, s_j\}$. We choose the tax rate, $\tau = 0.3$, the penalty rate, $T = 1.1$, and the probability of audit parameter, $\rho = 0.7$. To model an economy in which there are far fewer large firms than small firms, we use an exponential distribution with mean, $\mu = 0.8$ and truncated to $(0, 1]$. We model best response curves by having firm i choose c_i to maximize the expected profit from 100,000 transactions with partners drawn i.i.d. from the firm size distribution. The following figures show smoothed best response curves as one parameter is varied, and the tables show the likelihood of arriving at the all-formal equilibrium under each parameter value.

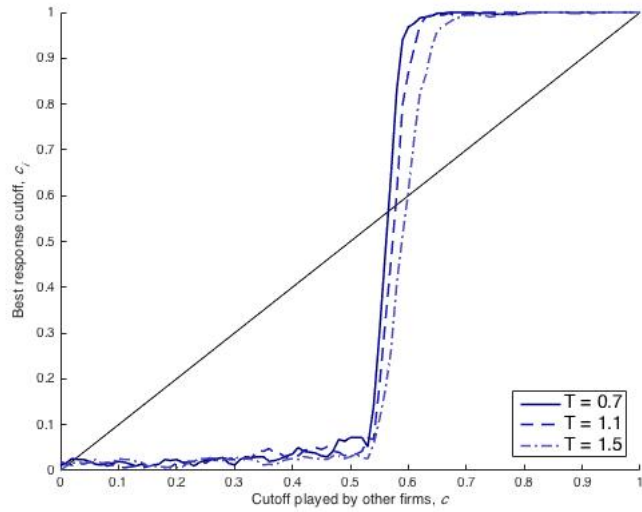
These simulations confirm the existence of a single, unstable intermediate equilibrium between the stable all-formal and all-informal equilibria and demonstrate that, under the parameters given, the best response curves are quite steep, suggesting rapid convergence toward one or the other equilibrium. The direction of change in the likelihood of converging to the all-formal equilibrium as predicted by propositions 3.6-3.9 is confirmed in each of our four simulations.

Figure 1: Best response curve varying tax rate (τ)



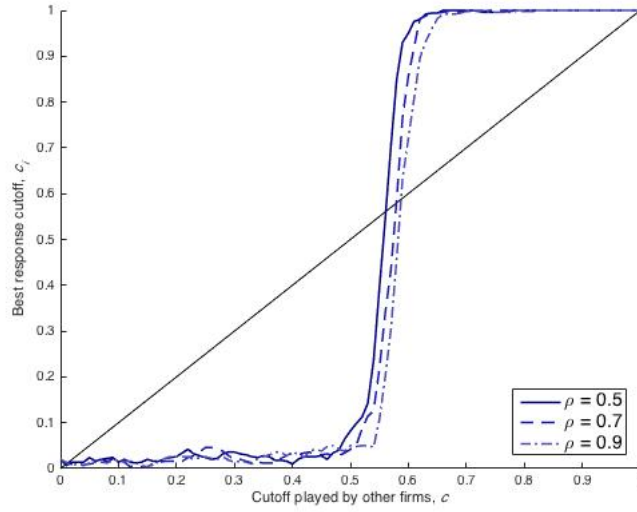
Parameter	Likelihood of convergence at all-formal equilibrium
$\tau = 0.15$	0.64
$\tau = 0.30$	0.58
$\tau = 0.45$	0.53

Figure 2: Best response curve varying penalty rate (T)



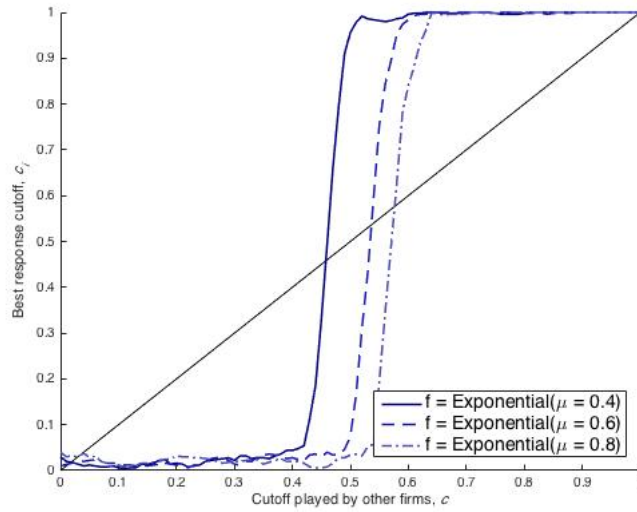
Parameter	Likelihood of convergence at all-formal equilibrium
$T = 0.7$	0.57
$T = 1.1$	0.58
$T = 1.5$	0.60

Figure 3: Best response curve varying likelihood of audit (ρ)



Parameter	Likelihood of convergence at all-formal equilibrium
$\rho = 0.5$	0.57
$\rho = 0.7$	0.58
$\rho = 0.9$	0.59

Figure 4: Best response curve varying distribution of firm sizes (f)



Parameter	Likelihood of convergence at all-formal equilibrium
$f = \text{Exp}(\mu = 0.4)$	0.46
$f = \text{Exp}(\mu = 0.6)$	0.54
$f = \text{Exp}(\mu = 0.8)$	0.58

4 Discussion

The model presented in the previous section yields the following conclusions: (1) formality is correlated with firm size; (2) informality increases when the tax rate increases, when the penalty rate for tax evasion decreases, or when the probability of audit decreases; and (3) informality decreases as the distribution of firm sizes shifts toward larger firms.

The first two results match stylized facts on tax compliance. We assess empirical support for each in the following subsections. These results are important in that the inter-firm mechanism which this paper brings to light agrees with existing theoretical and empirical models of tax compliance. The third conclusion, linking informality to the distribution of firm sizes, is perhaps this paper's most novel contribution. Several applications of this result, as well as empirical support for the model's predictions under those scenarios, are discussed in subsection 4.3.

4.1 Formality and firm size

There is a large body of empirical evidence to support the conclusion that informality is concentrated amongst small firms. Galiani and Weinschelbaum (2012) gather data on formality across twelve Latin American countries and find that this trend is true for all: the formality rate in Argentina, for instance, increases from 26.5% among small firms to 87.7% among large firms. Even when limiting the sample of firms considered to those registered with the government, Gatti and Honorati (2008) find that self-reported tax compliance increases linearly with firm size from 72% of sales declared in micro firms to an average of 86% in very large firms.

The original model of tax compliance developed by Allingham and Sandmo (1972) does not take firm size into account. Several models since, though, have done so in order to explain the correlation between formality and firm size. These models offer varying explanations including better access to credit, lower interest rates, and productive public goods that encourage larger firms to register formally (see, for instance, Straub (2005) and de Paula and Scheinkman (2006) as well as a survey of the broader literature by Galiani and Weinschelbaum (2012)). The model in this paper predicts that larger firms choose formality due to increased likelihood of audit. In a survey of informal firms across eight African nations, Amin and Huang (2014) find that large informal firms demonstrate a greater willingness to register formally than small informal firms. The authors note no difference between the perceived benefit by small and large informal firms of registering with regard to access to finance, raw materials, infrastructure, and government services and fewer bribes to pay, which leaves fear of penalty as a viable motivation for transition to formality with size.

4.2 The impact of shifting standard tax parameters: tax rate, penalty rate, and the likelihood of audit

The comparative statics predictions presented in this paper are also corroborated by a wide body of empirical evidence. Cebula (1997) confirms each of the models predictions em-

pirically for the federal income tax, finding that the size of the underground economy in the United States increases with the personal income tax rate and decreases with both the percentage of tax returns audited and the penalties imposed by the IRS on unpaid taxes.⁹ Tedds and Giles (2002) expand this result beyond the income tax, finding a strong positive correlation between the aggregate effective tax rate and the size of the underground economy in Canada, while data from Latin America provided by Loayza (1996) demonstrates that the correlation between the size of the informal sector and tax rates holds in developing nations as well. These findings are bolstered by experimental evidence from Kleven *et al* (2011). The study, conducted in conjunction with the Danish tax collection agency, shows that marginal tax rates have a positive effect on tax evasion and that threat-of-audit letters significantly increased tax payment on self-reported income.

A broader interpretation of our model’s tax rate parameter as a proxy for the burdens of registering formally and complying with regulation predicts that increases in the cost of registration and compliance result in an increase in the extent of informality. This prediction is substantiated by World Bank data on the size of the informal economy in African countries, which Verick (2006) shows is positively correlated with the time required to get a license, the cost of registering a property, and the number of documents required for exporting.

4.3 Changes in the distribution of firm sizes: Applications

Our model suggests that norms among small firms that are most prone to informality is influenced by the distribution of firm sizes in the economy. In particular, the existence of large firms—and the opportunity to transact with them—can encourage small firms on the margin to formalize and comply with taxes. We discuss four applications of this result: First, we examine the link between economic development and declining informality. Second, we apply our model in an explanation of the cross-industry variation in formality recorded by Artavanis, Morse and Tsoutsoura (2015) in Greece. We then proceed to discuss the impact of multinational entry and trade liberalization on informality.

4.3.1 Informality and development

La Porta and Shleifer (2008) find that the size of the informal economy falls with development, showing a strong negative correlation between income per capita and tax evasion. In a follow-up survey (2014), the authors assert that “as an economy develops, informality shrinks.” Of course, several confounding variables that emerge in the course of development complicate any single explanation of declining informality. Improved government institutions, adoption of paper trails that increase likelihood of detection, and an increase in the availability of low-cost credit all provide other explanations by which development may contribute to lower rates of informality, to name a few.

⁹While our model focuses on taxes paid by firms and not on tax compliance by individuals, empirical studies in the income tax literature are useful as they include the same standard tax parameters that are part of this model. Indeed, the way firms react to changes in the tax or penalty rate is unlikely to be substantially different from the way individuals react, and the studies cited here suggest this is the case in practice.

Industrialization entails a transition from an economy of small, agricultural firms to urban firms with larger economies of scale. Our model suggests that the emergence of large firms as a product of industrialization results in a greater extent of formality according to our model. This mechanism may be complementary to those discussed above whereby development and declining informality are linked.

4.3.2 Cross-industry variation in formality

Using data on household credit in Greece, Artavanis, Morse and Tsoutsoura (2015) estimate the level of tax evasion across industries. The authors find that tax evasion is greatest in professional services such as law and medicine and is lowest in personal services and manufacturing. Several parameters vary across these industries: the authors, for instance, find that tax evasion is most rampant in industries with a limited paper trail and in industries that have powerful government representation.

These findings are consistent with the predictions of this model. In particular, a limited paper trail and powerful government representation (which results in a lack of will to penalize tax evasion) reduce the threat of audit in certain industries, leading to an increase in the extent of informality as per our model. Distributions of firm sizes also differentiate industries: Law and medicine, found to be the primary tax-evading industries in the paper, usually involve small, boutique firms. In contrast, industries like manufacturing and retail involve large economies of scale and are thus populated by a greater number of large firms. The trend of greater informality amongst industries with firms that are smaller on average is evidenced by the paper's empirical findings, illustrating the relevance of our result in a cross-industry setting. Future cross-sectional studies that compare tax evasion at the firm level may prove especially useful in assessing this model's predictions and in providing guidance to more effective tax policy.

4.3.3 Multinational entry and the size of the informal economy

The impact of multinationals on the level of informality in countries they enter has been long disputed. Verick (2006), for instance, hypothesizes that recent growth in the size of the informal economy in sub-Saharan Africa is a result of multinationals who have an interest in keeping labor costs low. Schneider and Enste (2000), however, find that growth of the informal economy is not limited to developing nations: sizeable growth rates are found among Northern European countries and the United States as well. This certainly complicates a direct attribution of informality to multinational entry. Schneider and Enste (2000), for instance, instead attribute the increase in informal activity to the rise of the burden of tax, increased regulation, and declining tax morale.

Harsch (2001) reports that traditional activities of rural women in Burkina Faso were formalized when a number of multinational cosmetic companies began to purchase shea butter from locals. Such anecdotes suggest that trade opportunities with multinationals entering an economy can indeed encourage an increase in formality, which is the mechanism and prediction illustrated by the model in this paper. A body of international organization

literature, including Nadelmann (1990) and Wrage and Wrage (2005), also suggests that multinationals, subject to greater levels of scrutiny in their home countries, shift norms away from informality and corruption in the markets they enter. While these are useful cases and theories of how multinational entry influences norms of formality, empirical evidence to substantiate these arguments has yet to be presented.

4.3.4 The effects of trade liberalization on informality

A broader set of empirical evidence can be found in the literature (though limited) on informality and trade liberalization.¹⁰ The conventional view as articulated by Goldberg and Pavcnik (2003) holds that trade liberalization, like multinational entry, results in greater foreign competition and hence leads to an increase in the size of the informal sector. In an empirical study of the impact of NAFTA on informality in Mexico, however, Aleman-Castilla (2006) finds that reductions in Mexican import tariffs resulted in significant declines in the likelihood of informality in tradable industries. This result provides strong support for both the prediction and the mechanism given in this paper: our model attributes the decline in informality to trading opportunities with larger firms, which explains the concentration of informality declines among tradable industries in the NAFTA study. Fugazza and Fiess (2010) are more mixed in their conclusions: their study's cross-sectional correlations find that trade liberalization does indeed induce a reduction in informality, but the estimations of a model they develop find contrasting results across data sets.

5 Conclusion

As a contribution to the tax compliance literature, we have demonstrated a unique inter-firm mechanism that influences decisions on tax evasion. In doing so, we also offer theoretical predictions on how multinational entry and trade liberalization impact informality, questions which before now have largely been left unanswered. Where empirical literature is available, the predictions of this model are validated.

By necessity, this model focuses its attention on transactions between firms and, like all models, abstracts away several other factors that influence informality. Features of the labor market, in particular—including segregation of formal and informal workers by human capital or productivity—likely play an important role in shaping the level of informality, since the supply of labor to each sector of the economy defines the possible size of each. Attention, too, should be given to questions of political economy which lie between change of the parameters mentioned in this model and actual changes in the breadth of the tax base. Institutions and agents play a crucial role in both enabling and limiting informality, and while

¹⁰Some studies, such as Emran and Stiglitz (2003) and Boadway and Sato (2009), compare the efficacy of trade taxes and value-added taxes in the presence of an informal sector. While these studies are complementary to the findings in this paper, our model is focused on cases where trade liberalization makes trade with foreign firms possible that was prohibitively costly before liberalization. The conclusions of these studies may be taken hand in hand when considering how trade liberalization and tax policy design can impact tax collection.

blunt measures like effective tax rates or audit policies may shape firm decisions, ultimately human norms and relationships play the most important role in encouraging compliance.

The central motivation of this paper is to identify how the parameters and exogenous factors discussed here impact informality and thereby limit or enable the public investment that can spur development. While informality critically limits this investment, consequences of the discussed parameter changes on the labor market and on questions of political economy may have equally dramatic effects on the level and efficacy of public investment. To adequately assess the sum of these consequences requires not only an understanding of the channel on which this paper sheds light, but also on adjacent systems and often their specific, regional contexts.

With regard to the mechanism this paper describes, empirical work that assesses the importance of this channel among others developed in the literature would be useful. In particular, a cross-sectional study of formality by industry or a greater body of comparative studies on tax compliance across states may be especially useful in identifying the importance of this mechanism. We also identify open questions around the impact of multinational entry and trade liberalization on informality which have not received much attention in the literature. The theoretical predictions from this paper should be assessed by future investigations of the direction and magnitude of these effects in practice.

6 References

Acemoglu, D. and M. O. Jackson (2014). Social norms and the enforcement of laws. Working Paper. National Bureau of Economic Research.

Adreoni, J., B. Erard & J. Feinstein (1998). Tax compliance. *Journal of Economic Literature*, 36(2), 818-860.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4), 323-338.

Amin, M. and X. Huang (2014). Does firm-size matter in the informal sector? *World Bank Group Enterprise Notes*, No. 28.

Artavanis, N., A. Morse & M. Tsoutsoura (2015). Tax evasion across industries: Soft credit evidence from Greece. Working Paper. National Bureau of Economic Research.

Beron, K. J., H. V. Tauchen & A. D. Witte (1992). Tax compliance: An investigation using individual taxpayer compliance measurement program (TCMP) data. *Journal of Quantitative Criminology*, 9(2), 177-202.

Boadway, R. and M. Sato (2009). Optimal tax design and enforcement with an informal sector. *American Economic Journal: Economic Policy*, 1(1), 1-27.

- Böhme, M. H. and R. Thiele (2014). Informal-formal linkages and informal enterprise performance in urban West Africa. *European Journal of Development Research*, 26, 473-489.
- Cebula, R. J. (1997). An empirical analysis of the impact of government tax and auditing policies on the size of the underground economy: The case of the United States, 1973-94. *The American Journal of Economics and Sociology*, 56(2), 173-185.
- de Paula, A. and J. A. Scheinkman (2010). Value-added taxes, chain effects, and informality. *American Economic Journal, Macroeconomics*, 2(4), 195-221.
- Emran, M. S. and J. Stiglitz (2003). Price-neutral tax reform with an informal economy. In *NEUDC Conference*.
- Feinstein, J. S. (1991). An econometric analysis of income tax evasion and its detection. *The RAND Journal of Economics*, 14-35.
- Fugazza, M. and N. M. Fiess (2010). Trade liberalization and informality: new stylized facts. *UN Policy Issues in International Trade and Commodities Studies Series*, 43.
- Galiani, S. and F. Weinschelbaum (2012). Modeling informality formally: households and firms. *Economic Inquiry*, 50(3), 821-838.
- Gatti, R. and M. Honorati (2008). Informality among formal firms: Firm-level cross-country evidence on tax compliance and access to credit. *World Bank Policy Research Paper*, 4476.
- Goldberg, P. K. and N. Pavcnik (2003). The response of the informal sector to trade liberalization. *Journal of Development Economics*, 72(2), 463-496.
- Gordon, R. and W. Li (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, 93(7), 855-866.
- Harsch, E. (2001). Making trade work for poor women: Villagers in Burkina Faso find opening in the global market. *Africa Recovery*, 15(4), 6-7.
- Kaldor, N. (1963). Taxation for economic development. *The Journal of Modern African Studies*, 1(1), 7-23.
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pederson & E. Saez (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651-692.

- Kleven, H. J., C. T. Kreiner & E. Saez (2015). Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. Working paper. National Bureau of Economic Research.
- Kopczuk, W. and J. Slemrod (2006). Putting firms into optimal tax theory. *The American Economic Review*, 96(2), 130-134.
- Krugman, P. *et al.* (1992). Toward a counter-counterrevolution in development theory. *The World Bank Economic Review*, 6(1), 15-38.
- La Porta, R. and A. Shleifer (2008). The unofficial economy and economic development. Working paper 14520. National Bureau of Economic Research.
- La Porta, R. and A. Shleifer (2014). Informality and development. *Journal of Economic Perspectives*, 28(3), 109-126.
- Loayza, N. V. (1996). The economics of the informal sector: a simple model and some empirical evidence from Latin America. In *Carnegie-Rochester Conference Series on Public Policy*, 45, 129-152. North-Holland.
- Madisson, A. (2001). The world economy: A millennial perspective. Organization for Economic Cooperation and Development.
- Milgrom, P. and J. Roberts (1990). Rationalizability, learning and equilibrium in games with strategic complementarities. *Econometrica*, 58(6), 1255-1277.
- Nadelmann, E. A. (1990). Global prohibition regimes: The evolution of norms in international society. *International Organization*, 44(4), 479-526.
- Pomeranz, D. (2013). No taxation without information: Deterrence and self-enforcement in the value added tax. Working paper. National Bureau of Economic Research.
- Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, 643-663.
- Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38(1), 77-114.
- Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *The Journal of Economic Perspectives*, 21(1), 25-48.
- Slemrod, J., M. Blumenthal & C. Christian (2001). Taxpayer response to an increased

probability of audit: Evidence from a controller experiment in Minnesota. *Journal of Public Economics*, 79(3), 455-483.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In *Handbook of Public Economics*, 3, 1423-1470.

Straub, S. (2005). Informal sector: the credit market channel. *Journal of Development Economics*, 78(2), 299-321.

Tedds, L. M. and D. E. A. Giles (2002). *Taxes and the Canadian Underground Economy*. Canadian Tax Foundation.

Verick, S. (2006). The impact of globalization on the informal sector in Africa. Economic and Social Policy Division, *United States Economic Cat Commission for Africa*.

Wrage, S. and A. Wrage (2005). Multinational enterprises as moral entrepreneurs in a global prohibition regime against corruption. *International Studies Perspective*, 6, 316-324.

Zandt, T. V. and X. Vives (2007). Monotone equilibria in Bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1), 339-360.

Appendices

A Proofs

Proof of Proposition 3.1: Consider first the equilibrium in which all large firms are formal and all small firms are informal. For large firms:

$$u_i(a_i = 1; s_1) = \left(\frac{m-1}{N-1} \right) s_1^2 [1 - \tau]$$

$$u_i(a_i = 0; s_1) = \left(\frac{N-m}{N-1} \right) s_1 s_2 [1 - \rho T s_1]$$

Our condition for equilibrium is

$$\begin{aligned} \left(\frac{m-1}{N-1} \right) s_1^2 [1 - \tau] &\geq \left(\frac{N-m}{N-1} \right) s_1 s_2 [1 - \rho T s_1] \\ (1 - \tau) &\geq \frac{s_2 [1 - \rho T s_1]}{s_1} \left(\frac{N-m}{m-1} \right) \end{aligned} \tag{1}$$

For small firms:

$$u_i(a_i = 0; s_2) = \left(\frac{N - m - 1}{N - 1} \right) s_2^2 [1 - \rho T s_2]$$

$$u_i(a_i = 1; s_2) = \left(\frac{m}{N - 1} \right) s_1 s_2 [1 - \tau]$$

Our condition for equilibrium is

$$\left(\frac{N - m - 1}{N - 1} \right) s_2^2 [1 - \rho T s_2] \geq \left(\frac{m}{N - 1} \right) s_1 s_2 [1 - \tau]$$

$$(1 - \tau) \leq \frac{s_2 [1 - \rho T s_2]}{s_1} \left(\frac{N - m - 1}{m} \right) \quad (2)$$

Combining the conditions from (1) and (2), we have an equilibrium for any τ such that

$$\frac{s_2 [1 - \rho T s_1]}{s_1} \left(\frac{N - m}{m - 1} \right) \leq (1 - \tau) \leq \frac{s_2 [1 - \rho T s_2]}{s_1} \left(\frac{N - m - 1}{m} \right)$$

Now consider the equilibrium in which all large firms are informal and all small firms are formal. For large firms:

$$u_i(a_i = 0; s_1) = \left(\frac{m - 1}{N - 1} \right) s_1^2 [1 - \rho T s_1]$$

$$u_i(a_i = 1; s_1) = \left(\frac{N - m}{N - 1} \right) s_1 s_2 [1 - \tau]$$

Our condition for equilibrium is

$$\left(\frac{m - 1}{N - 1} \right) s_1^2 [1 - \rho T s_1] \geq \left(\frac{N - m}{N - 1} \right) s_1 s_2 [1 - \tau]$$

$$(1 - \tau) \leq \frac{s_1 [1 - \rho T s_1]}{s_2} \left(\frac{m - 1}{N - m} \right) \quad (3)$$

For small firms:

$$u_i(a_i = 1; s_2) = \left(\frac{N - m - 1}{N - 1} \right) s_2^2 [1 - \tau]$$

$$u_i(a_i = 0; s_2) = \left(\frac{m}{N - 1} \right) s_1 s_2 [1 - \rho T s_1]$$

Our condition for equilibrium is

$$\left(\frac{N - m - 1}{N - 1} \right) s_2^2 [1 - \tau] \geq \left(\frac{m}{N - 1} \right) s_1 s_2 [1 - \rho T s_1]$$

$$(1 - \tau) \geq \frac{s_1[1 - \rho T s_1]}{s_2} \left(\frac{m}{N - m - 1} \right) \quad (4)$$

Note that

$$\frac{m}{N - m - 1} > \frac{m}{N - m} > \frac{m - 1}{N - m}$$

Then,

$$\frac{s_1[1 - \rho T s_1]}{s_2} \left(\frac{m}{N - m - 1} \right) > \frac{s_1[1 - \rho T s_1]}{s_2} \left(\frac{m - 1}{N - m} \right)$$

so we find that there is no value of τ that satisfies both (3) and (4). This concludes the proof.

Proof of Lemma 3.2: Suppose firm i , with size s_i , weakly prefers transacting formally to transacting informally. Then,

$$E_j[a_j \pi(s_i, s_j)(1 - \tau)] - E_j[(1 - a_j) \pi(s_i, s_j)(1 - T \mu(s_i, s_j))] \geq 0$$

Substituting in our functional forms for π and μ ,

$$(1 - \tau) s_i E_j[a_j s_j] - s_i E_j[(1 - a_j)(s_j - T s_j \max\{s_i, s_j\})] \geq 0$$

Simplifying and dividing by s_i :

$$(1 - \tau) E_j[a_j s_j] - E_j[(1 - a_j) s_j] + T E_j[(1 - a_j) s_j \max\{s_i, s_j\}] \geq 0 \quad (5)$$

Note that

$$\begin{aligned} E_j[(1 - a_j) s_j \max\{s_i, s_j\}] &= \alpha (G(s_i) s_i E_{s_j|s_j < s_i, a_j=0}[s_j] + (1 - G(s_i)) E_{s_j|s_j > s_i, a_j=0}[s_j^2]) \\ &= \alpha \left(s_i \int_0^{s_i} s g(s) ds + \int_{s_i}^1 s^2 g(s) ds \right) \end{aligned}$$

where $G(s)$ is the cumulative distribution function of firm sizes for firms with $a_j = 0$ and α is the fraction of firms for which $a_j = 0$. The derivative of this expectation with respect to s_i is strictly positive

$$\frac{\partial}{\partial s_i} E_j[(1 - a_j) s_j \max\{s_i, s_j\}] = \alpha s_i^2 g(s_i) + \alpha \int_0^{s_i} s g(s) ds - \alpha s_i^2 g(s_i) = \alpha E_{s_j|s_j < s_i, a_j=0}[s_j] > 0$$

as long as a non-zero number of firms plays $a_j = 0$. So, we can conclude that, for some $s'_i > s_i$,

$$E_j[(1 - a_j) s_j \max\{s'_i, s_j\}] > E_j[(1 - a_j) s_j \max\{s_i, s_j\}] \quad (6)$$

To conclude the proof, we must show that a firm with size $s'_i > s_i$ strictly prefers transacting formally to transacting informally, i.e.

$$\Delta u(a, s'_i) = E_j[a_j \pi(s'_i, s_j)(1 - \tau)] - E_j[(1 - a_j) \pi(s'_i, s_j)(1 - T \mu(s'_i, s_j))] > 0$$

Substituting in our functional forms for π and μ into the left hand side of the above equation and simplifying, we get

$$\begin{aligned}\Delta u(a, s'_i) &= s'_i E_j[a_j s_j] - s'_i E_j[(1 - a_j) s_j] + s'_i T E_j[(1 - a_j) s_j \max\{s'_i, s_j\}] \\ &= s'_i (E_j[a_j s_j] - E_j[(1 - a_j) s_j] + T E_j[(1 - a_j) s_j \max\{s'_i, s_j\}]) \\ &> s'_i (E_j[a_j s_j] - E_j[(1 - a_j) s_j] + T E_j[(1 - a_j) s_j \max\{s_i, s_j\}]) \\ &\geq s'_i(0) = 0\end{aligned}$$

where step 3 follows from (5) and step 4 follows from (6). This concludes the proof.

Proof of Proposition 3.3: Given

$$c = \frac{1}{\rho T} \left(1 - (1 - \tau) \frac{(1 - F(c)) E[s_i | s_i > c]}{F(c) E[s_i | s_i < c]} \right)$$

the cutoff value, c , satisfies

$$F(c) E[s_j | s_j < c] (1 - \rho T c) = (1 - F(c)) E[s_j | s_j > c] (1 - \tau) \quad (7)$$

We show that the proposed equilibrium holds for all firms. For firms of size $s_i < c$:

$$E[u_i(a_i = 0; s_i)] = F(c) s_i E[s_j | s_j < c] (1 - \rho T \max\{s_i, E[s_j | s_j < c]\})$$

$$E[u_i(a_i = 1; s_i)] = (1 - F(c)) s_i E[s_j | s_j > c] (1 - \tau)$$

Our condition for equilibrium is

$$F(c) E[s_j | s_j < c] (1 - \rho T \max\{s_i, E[s_j | s_j < c]\}) \geq (1 - F(c)) E[s_j | s_j > c] (1 - \tau)$$

which holds from (7) since $\max\{s_i, E[s_j | s_j < c]\} \leq c$. For firms of size $s_i \geq c$:

$$E[u_i(a_i = 1; s_i)] = (1 - F(c)) s_i E[s_j | s_j > c] (1 - \tau)$$

$$E[u_i(a_i = 0; s_i)] = F(c) s_i E[s_j | s_j < c] (1 - \rho T s_i)$$

Our condition for equilibrium is

$$F(c) E[s_j | s_j < c] (1 - \rho T s_i) \leq (1 - F(c)) E[s_j | s_j > c] (1 - \tau)$$

which holds from (7) since $s_i \geq c$. So, we conclude that the equilibrium holds.

Proof of Lemma 3.4: Suppose that π can be written as $\pi(s_i, s_j) = \pi_1(s_i) \pi_2(s_j)$. Firm i , with size s_i , weakly prefers transacting formally to transacting informally. Then,

$$E_j[a_j \pi(s_i, s_j) (1 - \tau)] - E_j[(1 - a_j) \pi(s_i, s_j) (1 - T \mu(s_i, s_j))] \geq 0$$

$E_j[a_j\pi_1(s_i)\pi_2(s_j)](1 - \tau) - E_j[(1 - a_j)\pi_1(s_i)\pi_2(s_j)] + TE_j[(1 - a_j)\pi_1(s_i)\pi_2(s_j)\mu(s_i, s_j)] \geq 0$
Simplifying, and assuming $\pi_1(s_1) > 0$, we get

$$E_j[a_j\pi_2(s_j)](1 - \tau) - E_j[(1 - a_j)\pi_2(s_j)] + TE_j[(1 - a_j)\pi_2(s_j)\mu(s_i, s_j)] \geq 0 \quad (8)$$

We must show that a firm with size $s'_i > s_i$ strictly prefers transacting formally to transacting informally, i.e.

$$\Delta u(a, s'_i) = E_j[a_j\pi(s'_i, s_j)(1 - \tau)] - E_j[(1 - a_j)\pi(s'_i, s_j)(1 - T\mu(s'_i, s_j))] > 0$$

Substituting in our functional forms for π and μ into the left hand side of the above equation and simplifying, we get

$$\begin{aligned} \Delta u(a, s'_i) &= E_j[a_j\pi_1(s'_i)\pi_2(s_j)(1 - \tau)] - E_j[(1 - a_j)\pi_1(s'_i)\pi_2(s_j)] + TE_j[(1 - a_j)\pi_1(s'_i)\pi_2(s_j)\mu(s'_i, s_j)] \\ &= \pi_1(s'_i) (E_j[a_j\pi_2(s_j)(1 - \tau)] - E_j[(1 - a_j)\pi_2(s_j)] + TE_j[(1 - a_j)\pi_2(s_j)\mu(s'_i, s_j)]) \\ &> \pi_1(s'_i) (E_j[a_j\pi_2(s_j)(1 - \tau)] - E_j[(1 - a_j)\pi_2(s_j)] + TE_j[(1 - a_j)\pi_2(s_j)\mu(s_i, s_j)]) \\ &\geq \pi_1(s'_i)(0) = 0 \end{aligned}$$

where step 4 follows from (8) and step 3 holds if

$$\frac{\partial}{\partial s_i} E_j[(1 - a_j)\pi_2(s_j)\mu(s_i, s_j)] > 0$$

which holds when μ is strictly increasing in s_i and $\pi_2(s)$ is strictly positive.

Proof of Lemma 3.5: The game described is supermodular iff

1. c_i is a compact subset of \mathbb{R} ,
2. u_i is upper semi-continuous in c_i, c_{-i} , and
3. u_i has increasing differences in (c_i, c_{-i}) .

Given $c_i \in [0, 1]$ and the form of u_i , the first two conditions are satisfied. Condition (3) is satisfied if

$$\frac{\partial u_i}{\partial c_i \partial c} \geq 0$$

In expectation,

$$\begin{aligned} u_i(c_i) &= (1 - F(c_i))(1 - F(c))E_{s_i \geq c_i, s_j \geq c}[\pi(s_i, s_j)][1 - \tau] \\ &\quad + F(c_i)F(c)E_{s_i < c_i, s_j < c}[\pi(s_i, s_j)][1 - T\mu(s_i, s_j)] \end{aligned}$$

We find the derivative using truncated distributions:

$$\begin{aligned}
\frac{\partial u_i}{\partial c_i} &= \frac{\partial}{\partial c_i} \left[(1 - F(c_i))(1 - F(c))(1 - \tau) \frac{\int_{c_i}^1 \int_c^1 \pi(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2}{(1 - F(c_i))(1 - F(c))} \right. \\
&\quad \left. + F(c_i)F(c) \frac{\int_0^{c_i} \int_0^c \pi(s_1, s_2) (1 - T\mu(s_1, s_2)) f(s_1) f(s_2) ds_1 ds_2}{F(c_i)F(c)} \right] \\
&= \frac{\partial}{\partial c_i} \left[(1 - \tau) \int_{c_i}^1 \int_c^1 \pi(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2 + \int_0^{c_i} \int_0^c \pi(s_1, s_2) (1 - T\mu(s_1, s_2)) f(s_1) f(s_2) ds_1 ds_2 \right] \\
&= -(1 - \tau) \int_c^1 \pi(c_i, s_2) f(c_i) f(s_2) ds_2 + \int_0^c \pi(c_i, s_2) (1 - T\mu(c_i, s_2)) f(c_i) f(s_2) ds_2 \\
\\
\frac{\partial u_i}{\partial c_i \partial c} &= \frac{\partial}{\partial c} \left[-(1 - \tau) \int_c^1 \pi(c_i, s_2) f(c_i) f(s_2) ds_2 + \int_0^c \pi(c_i, s_2) (1 - T\mu(c_i, s_2)) f(c_i) f(s_2) ds_2 \right] \\
&= (1 - \tau) \pi(c_i, c) f(c_i) f(c) + \pi(c_i, c) (1 - T\mu(c_i, c)) f(c_i) f(c) \\
&\geq 0
\end{aligned}$$

where step 3 holds if $\pi(s_i, s_j) \geq 0$ and $\tau + T\mu(s_i, s_j) \leq 2$ for all s_i, s_j .

Proof of Proposition 3.6: From Milgrom and Roberts (1990), a supermodular game is indexed by parameter t if each player's payoff function is indexed by $t \in T$, some ordered set, and for all i , $u_i(c_i, c_{-i}, t)$ has increasing differences in (c_i, t) . We will show that our game is indexed by τ . The first condition holds since $\tau \in (0, 1)$, which is an ordered set. We can show increasing differences in (c_i, τ) by showing that

$$\frac{\partial u_i}{\partial \tau \partial c_i} \geq 0$$

In expectation,

$$\begin{aligned}
u_i(c_i) &= (1 - F(c_i))(1 - F(c)) E_{s_i \geq c_i, s_j \geq c} [\pi(s_i, s_j)] [1 - \tau] \\
&\quad + F(c_i)F(c) E_{s_i < c_i, s_j < c} [\pi(s_i, s_j)] [1 - T\mu(s_i, s_j)]
\end{aligned}$$

We find the derivative using truncated distributions:

$$\begin{aligned}
\frac{\partial u_i}{\partial \tau} &= \frac{\partial}{\partial \tau} \left[(1 - F(c_i))(1 - F(c))(1 - \tau) \frac{\int_{c_i}^1 \int_c^1 \pi(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2}{(1 - F(c_i))(1 - F(c))} \right] \\
&= - \int_{c_i}^1 \int_c^1 \pi(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial u_i}{\partial \tau \partial c_i} &= \frac{\partial}{\partial c_i} \left[- \int_{c_i}^1 \int_c^1 \pi(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2 \right] \\
&= \int_c^1 \pi(c_i, s_2) f(c_i) f(s_2) ds_2 \\
&= f(c_i) (1 - F(c)) E_{s_j > c} \pi(c_i, s_j) \\
&\geq 0
\end{aligned}$$

if $\pi(s_i, s_j) \geq 0$, $\forall s_i, s_j$. If a supermodular game is indexed by t , then an increase in t results in a monotone upward shift in the best response curve for each agent. An upward shift in the best response curve results in a leftward shift of the single, unstable equilibrium, and hence the likelihood of convergence at the all-formal equilibrium decreases.

Proof of Proposition 3.7: We show that the firms' payoff function, u_i , has increasing differences in $(c_i, -T)$ by showing that

$$\frac{\partial u_i}{\partial(-T) \partial c_i} \geq 0$$

In expectation,

$$u_i(c_i) = (1 - F(c_i))(1 - F(c)) E_{s_i \geq c_i, s_j \geq c} [\pi(s_i, s_j)] [1 - \tau] + F(c_i) F(c) E_{s_i < c_i, s_j < c} [\pi(s_i, s_j)] [1 - T \mu(s_i, s_j)]$$

We find the derivative using truncated distributions:

$$\begin{aligned}
\frac{\partial u_i}{\partial(-T)} &= \frac{\partial}{\partial(-T)} \left[\frac{F(c_i) F(c) \int_0^{c_i} \int_0^c \pi(s_1, s_2) (1 - T \mu(s_1, s_2)) f(s_1) f(s_2) ds_1 ds_2}{F(c_i) F(c)} \right] \\
&= \int_0^{c_i} \int_0^c \pi(s_1, s_2) \mu(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial u_i}{\partial(-T) \partial c_i} &= \frac{\partial}{\partial c_i} \int_0^{c_i} \int_0^c \pi(s_1, s_2) \mu(s_1, s_2) f(s_1) f(s_2) ds_1 ds_2 \\
&= \int_0^c \pi(c_i, s_2) \mu(c_i, s_2) f(c_i) f(s_2) ds_2 \\
&= f(c_i) F(c) E_{s_j < c} \pi(c_i, s_j) \mu(c_i, s_j) \\
&\geq 0
\end{aligned}$$

if $\pi(s_i, s_j) \mu(s_i, s_j) \geq 0$, $\forall s_i, s_j$.

Proof of Proposition 3.8: Let $T' = \alpha T$. Then we can represent the game under μ with $\alpha = 1$ and the game under μ' with $\alpha = \rho$. We conclude that the equilibrium cutoff c is decreasing in T' in a proof identical to that for Proposition 3.7.

Proof of Proposition 3.9: The main result from Zandt and Vives (2007) holds under the following conditions:

1. the utility function u_i is supermodular in a_i , has increasing differences in (a_i, a_{-i}) , and has increasing differences in (a_i, s) ; and
2. the beliefs mapping $p_i : T_i \rightarrow M_{-i}$ is increasing with respect to the partial order on M_{-i} of first-order stochastic dominance

Our utility function has increasing differences in (a_i, a_{-i}) since, as more firms switch to $a_j = 1$, firm i 's payoff from playing $a_i = 1$ increases. Assuming $\mu(s_i, s_j)$ is strictly increasing in its first argument, we also have increasing differences in (a_i, s) since firm i is encouraged to play $a_i = 1$ as the s_i —and therefore the likelihood of audit—increases as well as when the size of potential trading partners increases.

Condition (2) is weakly satisfied since all firms share a common prior. (Beliefs are affiliated.)

Applying the result from Zandt and Vives (2007), we find that best response curves are monotone in the distribution of firms sizes in the economy.

B Alternative formulation: Locked types

In Section 3.2, we introduce a dynamic version of the game: an initial norm is drawn randomly from which firms begin strategic play, eventually converging toward the all-formal or all-informal equilibrium. We can achieve the conclusions of propositions 3.6-3.9 in a version of the game with static play as well by introducing “locked” types. Suppose there is a small population of firms which are locked-in to either playing informally or formally. Switching formality for these firm types is prohibitively costly. The rest of the firms continue to play according to the payoff function given in Section 3.

If the number of locked firm types is large enough, the best response cutoff for any general firm shifts away from $c_i = 0$ when $c = 0$ and away from $c_i = 1$ when $c = 1$, since the other sector of the economy now contains potential partners. We now have internal equilibria that are no longer fixed at $c = 0$ and $c = 1$. From lemma 3.5, we have that these internal equilibria are stable. We can then repurpose the proofs of propositions 3.6-3.9 to conclude that these stable equilibria move in the same direction as our conclusions on the likelihood of convergence at the all-formal equilibrium.