

Formality, Informality, and Social Welfare*

John Bennett[†]

1 May 2008

Abstract

An industry is modeled in which entrepreneurs, who are heterogeneous in ability, may enter formally or informally. With a limited pool of entrepreneurs or an aggregate capital constraint, output supply is non-monotonic: for low and high prices only formality is chosen; at intermediate prices an informal-formal mix is chosen. Depending on the level of demand, informal supply may be a built-in stabilizer or destabilizer. Because some entrepreneurs choose informality, some workers forgo social benefits, but others gain employment; consumer surplus rises but aggregate profit may rise or fall. However, profit is redistributed towards less able entrepreneurs, and this may damage investment and growth.

JEL Classification: O17, D2

Keywords: formality, informality, social welfare.

*I am grateful to Saul Estrin for various helpful discussions and to the Leverhulme Trust for financial support.

[†]Department of Economics and Finance, and Centre for Economic Development and Institutions (CEDI), Brunel University.

Formality, Informality, and Social Welfare

1 Introduction

The literature on informality in developing economies has to a large extent focused on its employment aspect. This reflects the critical importance of informal employment, which comprises about 60% of the labour force in these economies. Though proportionately smaller than informal employment, comprising nearly 40% of GDP, the output from informal activity also plays a fundamental role, perhaps meriting more attention from researchers than it has so far received.¹ In this paper we examine how the pattern of informality/formality in an industry is related to the distribution of entrepreneurial abilities and the strength of the demand for the industry's output, as well as to various parameters relating to the relative costs and benefits affecting the entrepreneur's choice between formality and informality. We also consider welfare aspects of the industry equilibrium, an issue that seems to have been neglected in the literature. A message of the paper is that the informality is not necessarily 'bad'. In principle it can be welfare-improving, though the conditions for this are tightly defined.

Our analysis relates only to activities in which formality-versus-informality is an active choice for an entrepreneur. It is not meant to apply to the lower tier of informal activity, which has low value added, often only involving resale, with low capital-intensity and almost no employment. Rather, it relates to informal activity that can make profit, is more capital-intensive, involves employment and may

¹These figures, which are widely cited, are taken from Schneider and Enste (2000). A more recent figure given by Schneider (2006) is that the 'shadow economy,' defined as market-based legal production, constitutes 38.7% of GDP, on average, for 96 developing economies.

entail competition with formal firms (see Fields, 2005, on this distinction within the informal sector). This tier of informal activity can be seen as the equivalent of the small-scale entrepreneurial sector in developed economies (Maloney, 2004).

We assume that, unlike a firm with informal status, a formal firm must bear the cost of providing social benefits for workers. (Equivalently, it pays a minimum wage rate that does not apply to informal firms.) Also, it must pay the cost of registration (or a tax). However, there are two benefits from formality in the model: a formal firm obtains a productivity benefit from greater access to public goods and services, and it may be larger than a firm with informal status. If an informal firm were large it would attract attention and be penalized by the authorities. Given our other assumptions, if an entrepreneur chooses formality, he or she makes more profit if the firm is larger, so the option of being large is always taken up.² Specifically, we suppose that, for a given entrepreneur, a formal firm would use twice as much capital and labour as an informal firm would. However, the amount of labour an entrepreneur employs per unit of capital is assumed to depend on a parameter reflecting his or her entrepreneurial ability. Given the distribution of the ability parameter across the population of potential entrepreneurs, we determine the industry equilibrium in terms of which entrepreneurs enter and what status, formal or informal, is chosen by each of these.

We focus on cases in which there is some informality in equilibrium, finding that informal output supply may be non-monotonic in price. At relatively low prices there is only formal supply; at intermediate prices there is both formal and informal supply; while at high prices there is only formal supply. This result is found when

²ILO (2000) makes size part of its definition of informality. It defines an informal sector enterprise as private, unincorporated, unregistered and non-agricultural, with at least some output going to sale or barter, with size less than a threshold that is country specific.

there is a constraint on the supply of entrepreneurs. We consider two forms of the constraint: a limited pool of entrepreneurs and an aggregate capital constraint. For a uniform distribution of the ability parameter we thus show that as demand rises the proportion of the pool of entrepreneurs that chooses informality rises and then falls monotonically. We also obtain conditions under which informality acts as a built-in stabilizer or as a built-in destabilizer in the face of changing demand.

When deciding whether to enter formally, an entrepreneur does not internalize either the social benefits his or her firm provides to workers or the registration fee that is a transfer to the state. If they were internalized, formality would always be chosen. Given this non-internalization, we are in a second-best world. The option of informality causes some entrepreneurs to choose informality who would otherwise have chosen formality, and some to produce informally who would otherwise not have produced at all. The net result is that some workers forgo social benefit provision; but total employment rises, as does output and consumer surplus, while aggregate profit may rise or fall. Depending on the welfare weights used, the net effect on welfare in this static framework may be positive, but since an effect of informality is to redistribute profit from more to less able entrepreneurs, it may be damaging for investment and growth.³

In the literature analyzing informality, heterogeneity of entrepreneurial ability was introduced by Rauch (1991), using a variation of the Lucas (1978) model in which ability is represented by a multiplicative parameter on output. He thus explains why a ‘missing middle’ in the size distribution of firms obtains. Our

³Informality may also, in principle, have a dynamic benefit, as a possible stepping-stone to formality; that is it may allow an entrepreneur to test the market and his or her own ability without incurring sunk costs. Without such a stepping stone the entrepreneur may not enter. This argument is analyzed by Bennett and Estrin (2007).

formulation shares with Rauch the size-based definition of formality and the existence of a minimum wage rate (or social benefit provision) only in the formal sector, though we characterize entrepreneurial ability differently. However, at the cost of suppressing the endogenous choice of size for each firm that Rauch analyzes, we are able to examine in detail the product market and welfare ramifications.

Various models have now been developed using some of the Lucas-Rauch formulation of entrepreneurial ability (e.g., Fortin, Marceau and Savard, 1997, Amaral and Quintin, 2006, and de Paula and Scheinkman, 2006). Other building blocks of our model are also common in the literature, including the size differential associated with formal/informal status (e.g., Fortin, Marceau and Savard, 1997; Ranis and Stewart, 1999); higher wage rates paid by formal than by informal firms (e.g., Goldberg and Pavcnik, 2003; Banerji and Jain, 2007) and the productivity benefit of formality (e.g., Amaral and Quintin, 2006; Loayza and Rigolini, 2006). Factors that we do not include but which are also found in the literature include taxation (de Paula and Scheinkman, 2006), differential capital costs (Straub, 2005), bribery (Ihrig and Moe, 2004) and job matching (Bosch, 2007). One paper that, like ours, focuses on the product market is that of Banerji and Jain (2007). However, their focus is different, examining the endogenous differences in the quality of output between informal and formal firms.

In Section 2 the model is set up. Section 3 considers the implications of the model for the formal-informal mix. Section 4 examines welfare issues. Section 5 concludes. An appendix contains proofs.

2 The Model

Consider an industry producing a homogeneous good. A pool of entrepreneurs exists for this industry, each one of whom may choose to run a firm with informal status, or to run a firm with informal status, or to stay out of production. Running one firm requires an entrepreneur's full attention. A firm is a price-taker in the product market and in input markets. To allow for heterogeneity of entrepreneurial abilities we assume that a cost coefficient λ is attached to each entrepreneur, where $\lambda > 0$, and that λ varies across the pool of entrepreneurs. A higher value of λ reflects lower entrepreneurial ability. We refer to an entrepreneur with a given λ as a ' λ -entrepreneur.'

If a firm is informal it employs k units of capital and produces one unit of output. Setting the capital cost at unity, its capital cost is therefore k . In conjunction with the k units of capital, a λ -entrepreneur employs λ units of labour to produce the unit of output, thus incurring a wage cost of λw , where w is the wage rate. Hence, writing p for the unit price of output, with informality, a λ -entrepreneur earns profit

$$\pi^I = p - w\lambda - k. \quad (1)$$

Instead, any entrepreneur may choose formal status for a firm, in which case profit is

$$\pi^F = 2[\beta p - (w + s)\lambda - k] - c; \quad \beta > 1. \quad (2)$$

This equation, reflects two costs and two benefits compared to informality. c is a fixed cost of formality, which we refer to as a registration cost, though it may also be interpreted as a lump-sum tax. The second cost is that for each unit

of labour employed, in addition to the market wage, a firm with formal status must pay the amount s , which, for most of our analysis we refer to as the cost of providing social benefits; but we shall alternatively interpret $w+s$ as a government-imposed minimum wage for formal firms. Note that the λ -coefficient applies to an entrepreneur regardless of whether formal or informal status is adopted.

On the benefit side, it is assumed that, per package of labour and capital inputs, a formal firm earns βp of revenue, rather than the p earned by a formal firm. This may reflect that formal firms gain superior access to public goods such as public infrastructure, contract enforcement and property rights. It may also be the result of formal firms being able to sell their output to the government sector, fetching a higher price there than the market level p . The other assumed benefit is that formal status is associated with larger size. Given our other assumptions, an entrepreneur who adopts formality would always prefer larger to smaller size. To be specific, we assume that an entrepreneur who adopts formality uses twice as much capital and labour as when he or she adopts informality.⁴

We denote the opportunity cost for a λ -entrepreneur of running a firm in the industry we consider by $\bar{\pi}$. For simplicity, this is assumed to be independent of λ . This is consistent with entrepreneurial ability being specific to the industry, in which case, as in Rauch (1991), $\bar{\pi}$ could be the wage an entrepreneur could make as a member of the labour force. If, as measured by λ , entrepreneurial ability has value in other industries, it would be appropriate to make $\bar{\pi}$ increasing in λ .

We assume the industry is competitive and that there is free entry. Using

⁴Our assumptions are consistent with results of McKenzie and Sakho (2007) for Bolivia. They find that, the cost of registering as formal, in the form of then facing a liability for taxes may be more than outweighed by the increase in customer base that is associated with the firm being able to issue tax receipts.

(1) and (2), in Figure 1 we plot the loci for $\pi^I = \bar{\pi}$, $\pi^F = \bar{\pi}$, and $\pi^I = \pi^F$ in (λ, p) -space. These are

$$\pi^I = \bar{\pi}: \quad p(\lambda) = w\lambda + k + \bar{\pi} \equiv p^I(\lambda); \quad (3)$$

$$\pi^F = \bar{\pi}: \quad p(\lambda) = \frac{1}{\beta} \left[(w+s)\lambda + k + \frac{1}{2}(c + \bar{\pi}) \right] \equiv p^F(\lambda); \quad (4)$$

$$\pi^I = \pi^F: \quad p(\lambda) = \frac{1}{2\beta - 1} [(w + 2s)\lambda + k + c] \equiv \bar{p}(\lambda). \quad (5)$$

The relative heights and slopes of these loci depend on the signs of two inequalities.

Figure 1 is drawn on the assumptions that

$$2(\beta - 1)k + (2\beta - 1)\bar{\pi} > c \quad \text{and} \quad (6)$$

$$s - (\beta - 1)w > 0. \quad (7)$$

The intersection of all three loci occurs at $\lambda = \lambda^*$ and $p = p^*$, where

$$\lambda^* = \frac{2(\beta - 1)k + (2\beta - 1)\bar{\pi} - c}{2[s - (\beta - 1)w]}; \quad p^* = \frac{2sk + (2s + w)\bar{\pi} - cw}{2[s - (\beta - 1)w]}. \quad (8)$$

Given (6) and (7), $\lambda^* > 0$ and $p^* > 0$.⁵

(6) may not be unduly restrictive: it implies, for example, that the registration cost is less than the opportunity cost. (7), however, a more critical condition, implying that, for a λ -entrepreneur, $(w+s)\lambda/\beta$, the unit labour cost with formality, exceeds $w\lambda$, the unit labour cost with informality. We shall assume throughout

⁵If $\bar{\pi}$ is decreasing in λ (but $\bar{\pi} \geq 0$) there is no effect on the $\pi^I = \pi^F$ locus; and as λ rises the $\pi^F = \bar{\pi}$ and $\pi^I = \bar{\pi}$ loci have diminishing, but still positive slopes. Since, from (3)-(5), for $\lambda \geq \lambda^*$, the slope of $\pi^F = \bar{\pi}$ diminishes less quickly than the slope of $\pi^I = \bar{\pi}$ does, the dependence of $\bar{\pi}$ on λ cannot cause the loci to intersect a second time, and so there is no substantive effect on our analysis.

the text that (6) and (7) hold, leaving the main implications of one or both not holding to footnotes 7 and 8.

[Figure 1]

Consider the conditions that make (7) more likely to hold. One is if formal firms pay rents to workers. Teal (1996) finds strong evidence of rent sharing with workers by firms in formal manufacturing in Ghana, rents being significantly associated with the presence of unions, with public and (partial or total) foreign ownership, and with higher profits. He concludes that the 30% differential between formal and informal sectors suggested by Lewis (1954) is a substantial underestimate. It is also suggested by Collier and Gunning (1999) that in some African countries governments may generate rents for (formal) firms on the implicit understanding that the firms will pay high wage rates to workers. Thus, in our model the satisfaction of (7) may be linked to β exceeding unity. (7) is also more likely to hold if informal firms provide benefits that cost little but are greatly valued by workers. For example, informal firms may provide greater flexibility of hours, and family members may be employed without an explicit wage being paid (see World Bank, 2007, on such behaviour in Latin America). Finally, note that in their survey of evidence on African manufacturing Bigsten and Soderbom (2006) summarize evidence that wages are significantly greater in larger firms and that this can only partly be attributed to differences in worker characteristics.⁶

⁶For simplicity, we do not allow for the costs of capital being different for informal than for formal firms. Assuming that capital is more expensive for informal firms because of their lack of access to formal sources of finance, so that they use moneylenders, it becomes less likely that the amended version of (7) will be satisfied. However, informal firms may get the capital from family and friends. In Africa the interest rate for such loans is typically at or close to zero. For example, in Ghana, La Ferrara (2003) finds a real interest rate of approximately zero among kin groups.

Given (6) and (7), so that Figure 1 applies, informality may be chosen for $\lambda \geq \lambda^*$, but will never be chosen for $\lambda < \lambda^*$. Above (below) $\pi^I = \bar{\pi}$ in the figure, we have that $\pi^I(>) < \bar{\pi}$; above (below) $\pi^F = \bar{\pi}$, we have that $\pi^F(>) < \bar{\pi}$; and above (below) $\pi^I = \pi^F$, we have that $\pi^I(<) > \pi^F$. A horizontal line at any given price p enables us to see the range of λ -values associated with informality, formality, or staying out of the industry. The configuration of behaviour that obtains depends on the distribution of λ among the pool of entrepreneurs, and from this we can determine the output supply curve, which can be broken into informal and formal components.⁷ Combining this with the demand curve we can find what combination of informal and formal production obtains in equilibrium.

3 The Informal-Formal Mix

We adopt the following notation. $\pi^I = \pi^F$ for

$$\lambda(p) = \frac{(2\beta - 1)p - k - c}{w + 2s} \equiv \bar{\lambda}(p). \quad (9)$$

$\pi^I = \bar{\pi}$ for

$$\lambda(p) = \frac{p - k - \bar{\pi}}{w} \equiv \lambda^I(p), \quad (10)$$

⁷If one or both of (6) and (7) do not hold the three loci still intersect the p -axis at positive values of p . If (6) does not hold, but (7) holds, the loci do not intersect in the positive quadrant: the $\pi^I = \pi^F$ locus is above the $\pi^F = \bar{\pi}$ locus, which is above the $\pi^I = \bar{\pi}$ locus throughout the quadrant. If (6) holds but (7) does not, the intersection occurs in the positive quadrant but with $\pi^I = \bar{\pi}$ highest and $\pi^I = \pi^F$ lowest to the right of the intersection. If neither (6) nor (7) hold, there is no intersection in the positive quadrant, $\pi^I = \bar{\pi}$ being highest and $\pi^I = \pi^F$ lowest. In note 7 we consider for each case what the formal and informal supply curves look like for each of these cases.

and $\pi^F = \bar{\pi}$ for

$$\lambda(p) = \frac{\beta p - k - (c + \bar{\pi})/2}{w + s} \equiv \lambda^F(p). \quad (11)$$

We express aggregate supply and demand as per entrepreneur in the pool. The cumulative density of λ is denoted by $G(\lambda)$, and the supply of output by informal and by formal firms, per entrepreneur, by $q_s^I(p)$ and $q_s^F(p)$, respectively, where total supply per entrepreneur is $q_s(p) = q_s^I(p) + q_s^F(p)$. For brevity, we shall henceforth omit the phrase ‘per entrepreneur.’

Let λ_{\min} and λ_{\max} denote the respective lowest and highest values of λ in the pool. If only minimal entrepreneurial skills are required in the industry then λ_{\max} may be so large that, for practical purposes, we can treat it as infinite. A finite upper bound on λ may exist, however, because the entrepreneurial skills that are needed in the industry are relatively scarce, and this may be a binding constraint on the industry equilibrium. A similar effect will obtain even if the entrepreneurial skills are not scarce, but a complementary input is rationed. We shall consider the example of a binding constraint on the aggregate supply of capital below.

To keep the language simple we shall state our results on the assumption that the λ -distribution is continuous. The results can easily be written more generally.

Proposition 1 *Assume that $\lambda^* \in [\lambda_{\min}, \lambda_{\max})$. Then for $p \in [0, p^F(\lambda_{\min}))$ supply is zero. For $p \in [p^F(\lambda_{\min}), p^*]$ supply is only formal and is increasing in p . For $p \in (p^*, \bar{p}(\lambda_{\max})]$ supply is a mix of formality and informality and is increasing in p ; formal supply is increasing in p for all $p \in (p^*, \bar{p}(\lambda_{\max})]$, while informal supply increasing in p for $p \in (p^*, p^I(\lambda_{\max})]$, but decreasing in p for $p \in (p^I(\lambda_{\max}), \bar{p}(\lambda_{\max}))$. For $p \in [\bar{p}(\lambda_{\max}), \infty)$ supply is all formal and is constant.*

This proposition can be explained using Figure 1. Raising p from zero, formal

entry will become profitable for p above $p^F(\lambda_{\min})$ (where $\pi_F = \bar{\pi}$ cuts $\lambda = \lambda_{\min}$). For $p \in [p^F(\lambda_{\min}), p^*]$, there will be formal entry by all entrepreneurs with $\lambda \leq \lambda^F(p)$ (where p cuts the $\pi_F = \bar{\pi}$ locus). In this price range

$$q_s(p) = q_s^F(p) = 2\beta G[\lambda^F(p)].$$

For $p \in (p^*, \bar{p}(\lambda_{\max})]$, formal entry occurs by entrepreneurs with $\lambda \leq \bar{\lambda}(p)$, that is, where p cuts the $\pi^I = \pi^F$ locus. Informal entry occurs by entrepreneurs with $\bar{\lambda}(p) < \lambda \leq \min\{\lambda^I(p), \lambda_{\max}\}$, that is, between where p cuts the $\pi^I = \pi^F$ locus and where it cuts the $\pi^I = \bar{\pi}$ locus, but only up to $\lambda = \lambda_{\max}$. At $p = p^I(\lambda_{\max})$ the entire pool of entrepreneurs has entered and so the only effect of further increases in price is to induce switches from informality to formality. Thus, for the price range $p \in (p^*, \bar{p}(\lambda_{\max})]$,

$$q_s(p) = q_s^F(p) + q_s^I(p) = 2\beta G[\bar{\lambda}(p)] + G[\min\{\lambda^I(p), \lambda_{\max}\}] - G[\bar{\lambda}(p)]. \quad (12)$$

Total supply continues to rise until price reaches $\bar{p}(\lambda_{\max})$, when all entrepreneurs in the pool are producing formally. To summarize, above the minimum price that induces entry, low prices are associated with formal production only; intermediate prices are associated with a mix of formal and informal production; and high prices are associated with formal production only.

Intuitively, Proposition 1 can be explained in terms of the balance, for any λ and p , of the productivity benefit of formality, the greater labour costs of formality - especially so for higher λ - and the larger size (and so, potentially, ability to make profit) of a formal firm. Below $p = p^F(\lambda_{\min})$ all entry is unprofitable; but

if price exceeds $p^F(\lambda_{\min})$, though is still no greater than p^* , entry is profitable for an entrepreneur with sufficiently low λ , that is, sufficiently low labour costs. The impact of a low λ is greater for formality, which uses more labour, than for informality, and so, also taking into account the productivity benefit β of formality, formality is profitable for low enough λ . Given that (6) holds, the registration cost c of formality is not so great as to prevent this outcome.

When price is raised above p^* , both formality and informality are profitable for a range of λ -values. But, within this range, for entrepreneurs with relatively high λ , the greater labour costs of formality make it less profitable than informality. Thus, the entrepreneurs with a relatively high λ choose informality, while those with a relatively low λ can profitably bear the higher labour requirements of formality. As price is raised yet further, entrepreneurs with a higher λ can enter profitably, but for those near the higher end of this range, informality is more profitable than formality because the saving in labour costs outweighs the benefits of greater size and of the productivity benefit β . However, as price rises further, and more entrepreneurs choose formality, eventually all the pool of entrepreneurs chooses formality.

Let ϕ denote the proportion of active firms that is formal. The relationship between ϕ and p depends on the distribution of λ . Suppose the distribution is uniform with mean Λ and upper and lower bounds $\Lambda \pm \delta$ ($\Lambda > \delta > 0$), where $\Lambda - \delta < \lambda^* < \Lambda + \delta$. The cumulative density function is $G(\lambda) = (\lambda - \Lambda + \delta)/2\delta$ for $\lambda \in [\Lambda - \delta, \Lambda + \delta]$. Then we obtain the following.

Corollary 1 *For a uniform distribution of λ , $\lambda \in [\Lambda - \delta, \Lambda + \delta]$, where $\Lambda - \delta < \lambda^* < \Lambda + \delta$, the proportion ϕ of active firms that is formal is as follows: $\phi = 1$ for*

$p \in (p^F(\lambda_{\min}), p^*]$; $0 < \phi < 1$ with $d\phi/dp < 0$ for $p \in (p^*, p^I(\Lambda + \delta)]$; $0 < \phi < 1$ with $d\phi/dp > 0$ for $p \in (p^I(\Lambda + \delta), \bar{p}(\Lambda + \delta))$; and $\phi = 1$ for $p \in [\bar{p}(\Lambda + \delta), \infty)$.

As p is raised the proportion $1 - \phi$ of active firms that is informal increases monotonically from zero to a maximum (that is less than unity) at $p = p^I(\Lambda + \delta)$. It then falls monotonically, reaching zero at $p = \bar{p}(\Lambda + \delta)$. The supply curve in this case is illustrated in Figure 2. If informality were somehow ruled out as an option, the supply curve would be $\bar{q}_s^F(p)$, which includes the solid line extending below $p = p^*$ and the broken line that continues above $p = p^*$ until the entire pool of entrepreneurs is active, shown here as occurring at (q_2, p_2) , at which point the supply curve becomes vertical. Taking into account, however, that entrepreneurs have the option of formality, this option is exercised at p just above p^* . The aggregate supply curve $q_s^F(p) + q_s^I(p)$ is to the right of $\bar{q}_s^F(p)$ in this range. Since in this price range some entrepreneurs would choose formality if informality were not available, but choose informality when it is available, the formal supply curve $q_s^F(p)$ is to the left of $\bar{q}_s^F(p)$. As p rises through $p = p_1$, all entrepreneurs in the pool are producing. There is a steady switch of entrepreneurs from informality to formality, and this increases aggregate supply, but at a lower rate than for $p \in [p_1, p_2)$. When p reaches p_2 the entire pool of entrepreneurs is producing formally, with output at its maximum, $q = q_2$, and the supply curve becomes vertical.

[Figure 2]

Combining this characterization of the supply side with a demand curve for output, we have that if demand is sufficiently tight or sufficiently loose all supply will be formal; but for intermediate demand there will be a mix of formality and informality. However, if the assumption that $\lambda^* \in [\lambda_{\min}, \lambda_{\max})$ does not hold, the

non-monotonicity result disappears. If $\lambda_{\min} \geq \lambda^*$ then at the lowest price for which positive output occurs, some supply will be informal; and if price is raised far enough, informality will disappear. Alternatively, if $\lambda_{\max} < \lambda^*$, informality is never chosen. The non-monotonicity result also disappears if λ_{\max} is so large that, given that demand is not indefinitely large, market equilibrium is always to the left of the vertical portion of the supply curve.

A distinction may be drawn here. If the industry is such that a high level of entrepreneurial skills is needed, and these skills are in short supply relative to derived demand for them, then we shall have the case described in Proposition 1, and market equilibrium may well be on the vertical stretch of the supply curve, with no informality. If, however, only minimal entrepreneurial skills are required then λ_{\max} may be large, in which case we shall see some informality even if demand is at a relatively high level.⁸

However, non-monotonicity can result from constraints other than exhaustion of the pool of entrepreneurs. An aggregate capital constraint has similar effects to an upper bound on λ . Suppose there is a limit on the aggregate amount of capital available and that any rationing of capital is efficient in the sense that an entrepreneur with a lower λ gets priority over one with a higher λ . Assume that

⁸If one or both of (6) and (7) do not hold, the three loci in Figure 1 still cut the p -axis at positive values of p . If (6) does not hold, but (7) holds (so c , as well as s is relatively large) the loci do not intersect in the positive quadrant; but since their relative heights are the same as in the portion of Figure 1 to the north-east of $\{\lambda^*, p^*\}$, the behaviour that obtains is similar to that for entrepreneurs with $\lambda \geq \lambda^*$ in Figure 1. Thus, the main features of the analysis in the text obtain but with the exclusion of the result that for low enough λ and p only formality is chosen. If neither (6) nor (7) holds (c is relatively large but s is relatively small), then the loci are as depicted in Figure 1 with change that the locus labelled $\pi^F = \bar{\pi}$ in the figure is now $\pi^I = \bar{\pi}$. Informality is chosen for sufficiently low combinations of λ and p . Finally, if (6) holds but (7) does not (both s and c take relatively low values) there is no intersection in the positive quadrant and informality is never chosen.

so large that as p is raised the capital constraint binds first at price p_a , where $p_a > p^*$, i.e., where, in the absence of the capital constraint there would be both formal and informal production.

As p rises above p_a , some entrepreneurs wish to switch from informality to formality and others wish to enter informally. Since in equilibrium those preferring formality have a lower λ than those preferring informality, the constraint will bind for those preferring informality. The increased capital requirements of the entrepreneurs who wish to switch to formality will be satisfied, preventing additional entrepreneurs from entering informally and taking away capital from the least efficient of those who were already producing informally. The supply of formal output will rise by the same amount as it would in the absence of the capital constraint; informal supply will fall; and aggregate supply will rise, but by less than it would in the absence of the capital constraint.

As p rises further, this process will continue until $p = p_b$ is reached, at which all informal supply disappears, with all the supply of capital being used by firms with formal status. For any rises in p above p_b the capital constraint prevents more entrepreneurs from entering production. Thus, for $p \geq p_b$ there is no informal supply and the aggregate (formal) supply curve is vertical. This implies the following.

Corollary 2 *If the distribution of λ is unbounded from above, an aggregate capital supply constraint that binds for some $p \in (p^*, \infty)$ will cause a non-monotonic response of informal supply to variation of price.*

3.1 Comparative Statics

From (8) we find that both λ^* and p^* are increasing in w , k , β , and $\bar{\pi}$, and decreasing in s and c . An increase in s can be interpreted as a rise in the minimum wage; but also, an increase w together with a decrease in s by the same amount is equivalent to a rise in the formal wage rate with the formal wage rate held constant. Thus, λ^* and p^* are decreasing in the minimum wage rate and in the informal wage rate.

Consider the implications of an increase in s from s' to s'' ; denote the corresponding values of λ^* and p^* by $\lambda^{*'}$ and $p^{*'}$ and $\lambda^{*''}$ and $p^{*''}$, respectively.

Lemma 1 *For an increase in social benefits from s' to s'' (or, equivalently, an increase of the minimum wage) $\lambda^{*'}$ > $\lambda^{*''}$ and $p^{*'}$ > $p^{*''}$. For $p \in [p^F(\lambda_{\min}), p^{*''}]$ entrepreneurs with $\lambda \in [(\beta p - k - (c + \bar{\pi})/2)/(w + s''), (\beta p - k - (c + \bar{\pi})/2)/(w + s')]$ will no longer produce. For $p \in (p^{*''}, \infty)$ entrepreneurs with $\lambda \in (((2\beta - 1)p - k - c)/(w + 2s''), ((2\beta - 1)p - k - c)/(w + 2s'))$ will switch from formality to informality. The number of firms and output per producing firm both fall.*

This lemma is illustrated in Figure 3, which reproduces the three loci from Figure 1 (these apply for $s = s'$) and adds, with broken lines, the new loci for $s = s''$. For any given λ , a higher p is required for a formal firm to achieve a profit of $\bar{\pi}$, and since entrepreneurs with higher λ have greater labour costs, the required increment to p is greater for higher λ . Thus, the increase in s rotates the $\pi^F = \bar{\pi}$ locus anti-clockwise around the point at which it meets the p -axis. Since π^I is unaffected by the change in s , a similar argument applies with respect to the $\pi^I = \pi^F$ locus, which also rotates anti-clockwise. However, the $\pi^I = \bar{\pi}$ locus

is unaffected. The intersection of $\pi^F = \bar{\pi}$ and $\pi^I = \pi^F$ slides leftward down the $\pi^I = \bar{\pi}$ locus to $(\lambda^{*''}, p^{*''})$, which is denoted by b in the figure.

[Figure 3]

The results in the lemma follow from our previous discussion. At $\lambda = \lambda_{\min}$ the rise in the $\pi^F = \bar{\pi}$ locus implies that a slightly higher price is required for a λ_{\min} -entrepreneur to produce formally. A similar argument applies for all $\{\lambda, p\}$ -combinations in the portion of the area abc to the right of $\lambda = \lambda_{\min}$. For $\lambda > \lambda^{*''}$, lower bound on $\{\lambda, p\}$ -combinations for which informality was originally profitable ($\pi^I > \bar{\pi}$) is unaffected by the change in s . Above this bound, for $\{\lambda, p\}$ -combinations at which $\pi^F > \pi^I$ by a sufficiently small amount for $s = s'$, then $\pi^F < \pi^I$ for $s = s''$, so that the entrepreneurs concerned switch from formality to informality. These $\{\lambda, p\}$ -combinations belong to the area bounded by bc and the $\pi^I = \pi^F$ loci before and after the change in s . With fewer firms producing at $s = s''$ than at $s = s'$ and with a larger proportion being formal, both total output and output per producing firm falls.

Similar considerations apply for variation of other parameter values. We omit description here because of space considerations. The effects can be found using the equations for the three loci to determine how the relationships depicted in Figure 1 must be amended for any parameter change. However, we note here how the model can be related to the cyclical behavior of the economy.

Again assuming a uniform distribution of labour costs, this can be seen by going to Figure 2. Suppose first that the demand curve cuts both $q_s^F(p)$ and $q_s^F(p) + q_s^I(p)$ in the range $p \in (p_1, p_2)$, that is, in the upper portion of the quadrilateral. Then it also cuts the broken portion of the $\bar{q}_s^F(p)$ -line in this range. As can be seen

from the relative slopes, a small vertical shift downwards in demand would cause a bigger fall in output if informality were prohibited than if it were not. In this sense informality is a built-in stabilizer. However, if the intersections with the demand curve occur in the range $p \in (p^*, p_1)$, that is, in the lower portion of the quadrilateral, the supply curve with informality is flatter than without formality. In this range the existence of informality adds to output instability.

This argument relates to output, but it can be reformulated in terms of employment. For $p \in (p^*, p_1)$ we again find that informality adds to instability; but the result changes for $p \in (p_1, p_2)$. In this price range, if informality were unavailable to entrepreneurs, formal employment would rise with p . If, however, informality is an option, since all entrepreneurs in the pool are producing in this price range, variation of p causes only a shift between formality and informality. For each λ -entrepreneur that shifts from informality to formality λ informal jobs are lost and 2λ formal jobs are created. Therefore, as p rises in this range employment falls; that is, in terms of employment, informality is a stronger built-in stabilizer than it is for production.⁹

4 Welfare

To examine the impact of informality on the components of welfare, we assume that the equilibrium price exceeds p^* , so that both informality and formality occur in equilibrium. We assume linear demand, $p = a - q_d$, where a is a positive constant and q_d the quantity demanded and a uniform distribution of λ , with $\lambda_{\min} = \Lambda - \delta < \lambda^*$ and $\lambda_{\max} = \Lambda + \delta \geq \lambda^I(p^{IF})$, where p^{IF} is the equilibrium price.

⁹Loayza and Rigolini (2006) find that in the short run, informal employment is counter-cyclical for the majority, but not all, developing countries.

This implies that, in terms of Figure 2, the equilibrium is in the range $p \in [p^*, p']$; that is, informal (and formal) supply are increasing in p (there are no restrictions on informality caused by the upper bound λ_{\max}).

Suppose that informality is ruled out. Then, using our earlier notation, supply is $\bar{q}_s^F(p)$, and we denote the equilibrium price and quantity by (\bar{p}^F, \bar{q}^F) . First we note that if, starting with the situation in which informality is ruled out, if p is fixed at \bar{p}^F , the introduction of the option of informality raises the supply of output.

Lemma 2 *Holding price constant at \bar{p}^F , the introduction of the option of informality raises the supply of output by*

$$\Delta q = \frac{(\bar{p}^F - p^*)z^2}{\delta w(w + s)(w + 2s)}. \quad (13)$$

where $z = s - (\beta - 1)w$.

Consider the impact of the introduction of informality as an option. Some entrepreneurs then switch to informality and some additional entrepreneurs who would not otherwise have produced, produce informally. The quantity supplied formally falls and that supplied informally rises, but as shown by the lemma, the total quantity supplied increases. Thus, the total supply curve with informality $q_s^F(p) + q_s^I(p)$ is to the right of $\bar{q}_s^F(p)$, the supply curve without informality, for $p > p^*$. Demand and supply for this case are depicted in Figure 4. With informality, the equilibrium price and quantity are (p^{IF}, q^{IF}) , with q^F being supplied formally and $q^{IF} - q^F$ informally. If, however, the option of informality were not open to firms, the formal supply curve would be $\bar{q}_F^s(p)$ and the equilibrium price and quantity would be (\bar{p}^F, \bar{q}^F) . Given Lemma 2, $\bar{p}^F > p^{IF} > p^*$ and $q^F < \bar{q}^F < q^{IF}$.

[Figure 4]

We can now examine the impact of the option of informality on social benefits, registration fees, consumer surplus and profits. Note that all the entrepreneurs producing formally when informality is an option also produce formally when it is not. Therefore, since formal output q^F with informality is smaller than formal output \bar{q}^F without informality, the provision of social benefits is smaller when informality is an option. The same conclusion follows for registration fees. However, since $\bar{p}^F > p^{IF}$ and $\bar{q}^F < q^{IF}$, consumer surplus is greater when informality is an option. The remaining component of welfare is profits, to which we now turn.

To compare aggregate profits with and without the option of informality, each equilibrium value of p must be determined and fed back into the profit equations. But this leads to intractable complications, and so we focus on simple cases. Suppose that $p = a = \bar{p}^F$, regardless of quantity. Using (2), (9), (11) and (8), the option of informality changes aggregate formal profit by $\Delta\pi^F = - \int_{\bar{\lambda}(\bar{p}^F)}^{\lambda^F(\bar{p}^F)} 2[\beta\bar{p}^F - \lambda(w + s) - k - c/2]d\lambda < 0$. $\Delta\pi^F < 0$ because some entrepreneurs who would choose informality when it is available would choose formality if informality were excluded. Using (1) and (10), aggregate informal profit rises from zero to $\Delta\pi^I = \int_{\bar{\lambda}(\bar{p}^F)}^{\lambda^I(\bar{p}^F)} (\bar{p}^F - \lambda w - k)d\lambda > 0$. Hence,

$$\Delta\pi^F + \Delta\pi^I = \frac{[(2s + w)\bar{\pi} + (\bar{p}^F - p^*)z](\bar{p}^F - p^*)z}{(2s + w)(s + w)w}. \quad (14)$$

Since $\bar{p}^F > p^*$ by assumption, $\Delta\pi^F + \Delta\pi^I > 0$ in this case. For a downward-sloping demand curve a simple condition cannot be found specifying when $\Delta\pi^F + \Delta\pi^I$ is positive or negative, but numerical examples confirm that in the extreme case of

the demand curve becoming vertical, $\Delta\pi^F + \Delta\pi^I$ may be negative.¹⁰ Thus, the impact of informality on profits can be of either sign.

Proposition 2 *Assume that λ is uniform with $\lambda \in [\Lambda - \delta, \Lambda + \delta]$, where $\Lambda - \delta < \lambda^*$ and $\Lambda + \delta$ is at least as large as $\lambda^I(\bar{p}^F)$, that demand is linear and that in the absence of informality $p > p^*$. Then the introduction of the option of informality damages some workers (it reduces social benefit provision) but benefits other workers (aggregate employment rises). It benefits consumers (consumer surplus expands) and may benefit employers (aggregate profit may rise).*

The impact of informality is to redistribute welfare. There is a reduction in social benefit provision, but an increase in consumer surplus and employment, and there may be an increase in aggregate profit. Entrepreneurs are brought into production who are less able than those who produce when informality is ruled out. Also, some entrepreneurs produce informally who would otherwise produce formally, thereby lowering their productivity. Thus, there is a redistribution of profit away from more able entrepreneurs towards less able ones and the average productivity of labour and capital is reduced. In a dynamic context this might have adverse consequences for investment and growth.

5 Conclusion

We have characterized an industry in which entrepreneurs may enter formally or informally. If a constraint exists on the amount of entry, supply is non-monotonic in

¹⁰If, for example, $k = c = \bar{\pi} = 0$, $w = 1$, $s = 0.5$, $\beta = 1.1$, $\Lambda - \delta = 0.1$ and $\bar{p}^F = 1$, then the introduction of informality as an option reduces price to approximately 0.938 and $\Delta\pi^F + \Delta\pi^I \approx -0.015$.

that for low and high prices only formality is chosen, while at intermediate prices some entrepreneurs choose formality and other, less able, entrepreneurs, choose informality. We consider two such constraints - a limited pool of entrepreneurs and an aggregate capital constraint. In a more general model other constraints such as a shortage of skilled labour or of intermediate goods might play this role. As we might expect intuitively, an increase in the social benefit provision required of formal firms, or, equivalently, an increase in the minimum wage that applies only to these firms, causes a contraction in output and in output per firm. However, when demand varies, informality may not be a built-in stabilizer - it may exacerbate instability.

The existence of the informal option for firms has a redistributive welfare effect. Social benefits are forgone by some workers, employment rises; consumer surplus rises, while aggregate profit may rise or fall. Depending on the weights in the welfare function informality may, in principle, be welfare enhancing. Thus, in a static model it is not necessarily bad to have ‘unfair competition’ by firms who do not pay social benefits or the minimum wage. However, since informality has the effect of redistributing production and profit production and profit towards less able entrepreneurs, investment and growth may be damaged.

Our analysis is based on the assumption that differences in entrepreneurial ability are manifested in terms of differing labour costs. Heterogeneity might instead have been represented by a multiplicative coefficient on output that differs across entrepreneurs. Thus, we would have $\pi^I = \alpha p - w - k$ and $\pi^F = 2(\alpha\beta p - w - s - k) - c$, where $\alpha > 0$ and a higher α denotes a more able entrepreneur. A necessary condition for informality to be chosen is then that $s > (\beta - 1)(w + k) + [(2\beta - 1)\bar{\pi} - c]/2$. If this condition is satisfied, informality is chosen for

$\alpha \in [(w + k + \bar{\pi})/p, (w + 2s + k + c)/(2\beta - 1)p)$ and formality is chosen for $\alpha \in [(w + 2s + k + c)/(2\beta - 1)p, \infty)$. As p is raised informal supply falls, but formal supply increases such that total supply rises. The analysis is then similar to our analysis above, except that a non-monotonicity result is not obtained.¹¹

References

- [1] Amaral, Pedro S, and Erwan Quintin (2006), A competitive model of the informal sector, *Journal of Monetary Economics*, 53, 1541-53.
- [2] Banerji, Arup, and Sanjay Jain (2007), Quality dualism, *Journal of Development Economics*, 84, 234-250.
- [3] Bennett, John, and Saul Estrin (2007), Informality as a stepping-stone: entrepreneurial entry in a developing economy, IZA Discussion Paper No 2950.
- [4] Bigsten, Arne, and Mans Soderbom (2006), What have we learned from a decade of manufacturing surveys in Africa? *World Bank Research Observer* 21, 241-265.
- [5] Bosch, Mariano (2007), Job creation and job destruction in the presence of informal labour markets, London School of Economics, mimeo.
- [6] Collier, Paul, and Jan Willem Gunning (1999), Explaining African growth performance, *Journal of Economic Literature*, 37, 12, March, 64-111.

¹¹Corresponding to Figure 1, a figure can be drawn in (α, p) -space. The curves (in the positive quadrant) for $\pi_I = \pi_F$, $\pi_I = \bar{\pi}$ and $\pi_F = \bar{\pi}$ are downward-sloping and non-intersecting. For $s > (\beta - 1)(w + k) + [(2\beta - 1)\bar{\pi} - c]/2$, $\pi_I = \pi_F$ is the highest and $\pi_I = \bar{\pi}$ the lowest.

- [7] Fields, Gary (2005), A guide to multi-sector labor market models, Cornell University, mimeo.
- [8] Fortin, Bernard, Nicolas Marceau and Luc Savard (1997), Taxation, wage controls and the informal sector, *Journal of Public Economics*, 66, 293-312.
- [9] Goldberg, Nina, and Pinelopi Pavcnik (2003), The response of the informal sector to trade liberalization, *Journal of Development Economics*, 72, 463-496.
- [10] Ihrig, Jane, and Karine S. Moe (2004), Lurking in the shadows: the informal sector and government policy, *Journal of Development Economics*, 73, 541-557.
- [11] International Labour Office (2000), Resolution concerning statistics of employment in the informal sector, in *Current International Recommendations on Labor Statistics*, 2000 edition, ILO, Geneva.
- [12] La Ferrara, Eliana (2003), Kin groups and reciprocity: a model of credit transactions in Ghana, *American Economic Review*, 93, 1730-1751.
- [13] Levy, Santiago (2007), Can social programs reduce productivity and growth? A hypothesis for Mexico, mimeo.
- [14] Lewis, W. Arthur (1954), Economic development with unlimited supplies of labour, *Manchester School*, 28, 139-191.
- [15] Loayza, Norman V., and Jamele Rigolini (2006), Informality trends and cycles, World Bank Policy Working Paper 4078.
- [16] Lucas, Robert E. (1978), On the size distribution of business firms, *Bell Journal of Economics*, 9, 508-523.

- [17] Maloney W. (2004), Informality revisited, *World Development*, 32, 1159-1178.
- [18] McKenzie, David, and Yaye Seynabou Sakho (2007), Does it pay firms to register for taxes? The impact of formality on firm profitability. IZA Discussion Paper No 3179.
- [19] Paula, Aureo de and Jose Scheinkman (2006), The informal sector, UCLA Department of Economics Working Paper.
- [20] Ranis, Gustav, and Frances Stewart (1999), V-goods and the role of the urban informal sector in development, *Journal of Economic Development and Cultural Change*, 47, 259-288.
- [21] Rauch, James E. (1991), Modelling the informal sector informally, *Journal of Development Economics*, 35, 33-47.
- [22] Schneider, Friedrich (2006), Shadow economies of 145 countries all over the world: what do we really know? Working Paper, Johannes Kepler University of Linz.
- [23] Schneider, Friedrich, and Dominik Enste (2000), Shadow economies: size, causes, and consequences, *Journal of Economic Literature*, 38, 77-114.
- [24] Straub, Stephane (2005), Informal sector: the credit market channel, *Journal of Development Economics*, 78, 299-321.
- [25] Teal, Frances (1996), The size and sources of economic rents in a developing country manufacturing labour market, *Economic Journal*, 106, 963-976.
- [26] World Bank (2007) Informality: Exit and Exclusion, World Bank, Washington D.C.

Appendix: Proofs

Corollary 1

This follows immediately from Proposition 1 except for $p \in (p^*, p^I(\Lambda + \delta)]$. In this range, $\phi = G[\bar{\lambda}(p)]/G[\lambda^I(p)] = [\bar{\lambda}(p) - \Lambda + \delta]/[\lambda^I(p) - \Lambda + \delta]$. Using (9)-(10), $\phi = w[p(2\beta - 1) - k - c - 2s\Lambda + 2s\delta - w\Lambda + w\delta]/(2s + w)(\delta w - \Lambda w + p - k - \bar{\pi})$. Thus, $d\phi/dp = vw/(2s + w)(p - k - \bar{\pi} - w\Lambda + w\delta)^2$, where $v = 2(\Lambda - \delta)(s - (\beta - 1)w) + c - (2\beta - 1)\bar{\pi} + 2k(1 - \beta)$ and so $sign(d\phi/dp) = sign v$. But $\Lambda - \delta < \lambda^*$ by assumption, so that, using (8), $v < 0$. Hence $d\phi/dp < 0$.

Lemma 2

At $p = \bar{p}^F$, if informality is ruled out, formal output is supplied by all entrepreneurs with $\lambda \leq \lambda^F(\bar{p}^F)$; i.e., $q^F(\bar{p}^F) = [\lambda^F(\bar{p}^F) - \Lambda + \delta]\beta/\delta$. If informality is an option, formal output is supplied by all entrepreneurs with $\lambda \leq \bar{\lambda}(\bar{p}^F)$; i.e., $q^F(\bar{p}^F) = [\bar{\lambda}(\bar{p}^F) - \Lambda + \delta]\beta/\delta$. Using (9) and (11), the fall in formal output supply from the introduction of the option of informality is therefore found to be $\varsigma = \beta(\bar{p}^F z - sk - \bar{\pi}s + cw/2 - \bar{\pi}w/2)/\delta(w + s)(w + 2s)$. From (8), $sk + (2s + w)\bar{\pi}/2 - cw/2 = zp^*$, so that $\varsigma = \beta z(\bar{p}^F - p^*)/\delta(w + s)(w + 2s)$. Informal output is supplied by entrepreneurs for whom $\lambda \in (\bar{\lambda}(\bar{p}^F), \lambda^I(\bar{p}^F)]$, i.e., supply is $q^I(\bar{p}^F) = [\lambda^I(\bar{p}^F) - \Lambda + \delta]/2\delta - [\bar{\lambda}(\bar{p}^F) - \Lambda + \delta]/2\delta = [\lambda^I(\bar{p}^F) - \bar{\lambda}(\bar{p}^F)]/2\delta$. Using (9), (10) and (8), this is found to equal $z(\bar{p}^F - p^*)/\delta w(w + 2s)$. (13) follows.

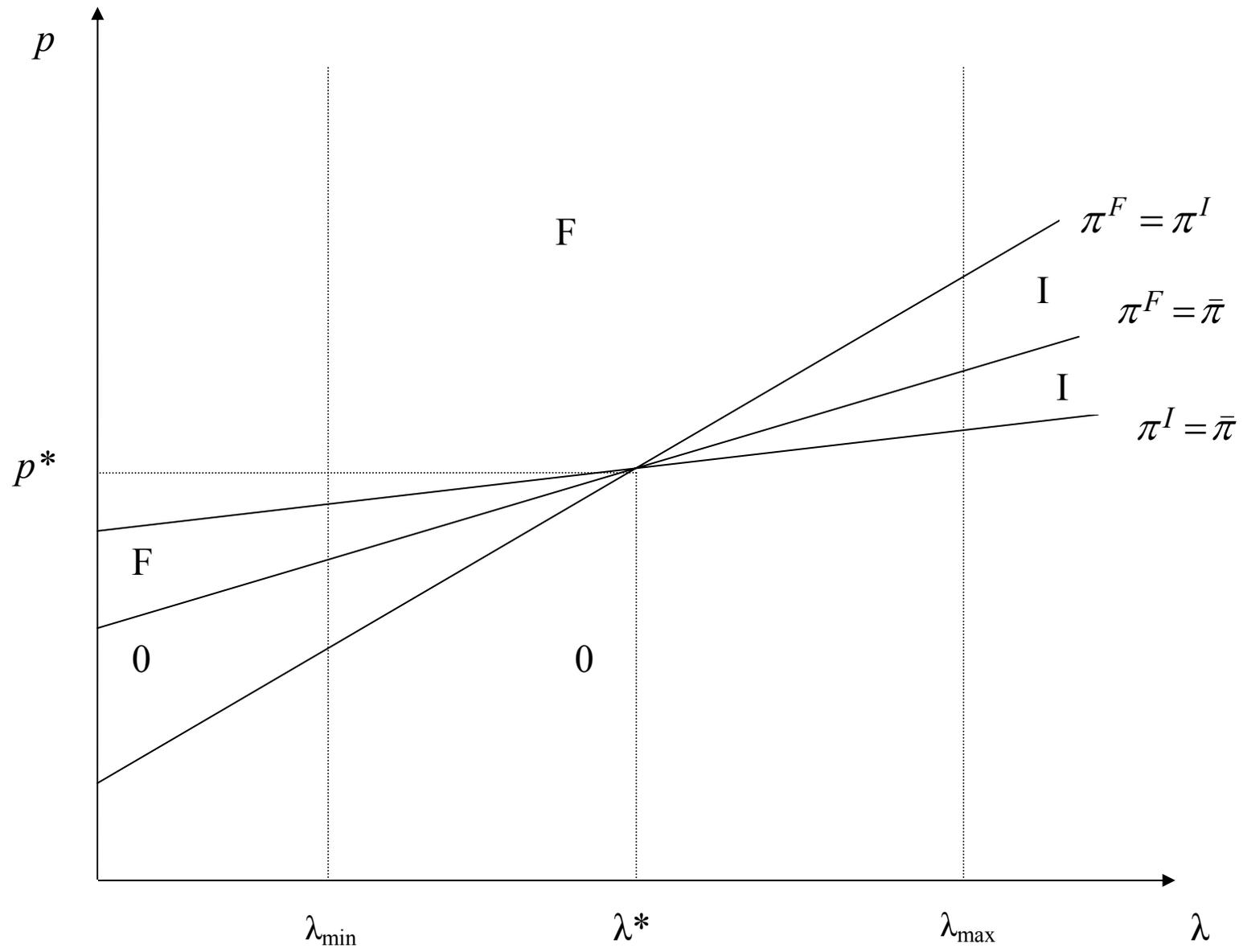


Fig. 1 The Informal-Formal Mix

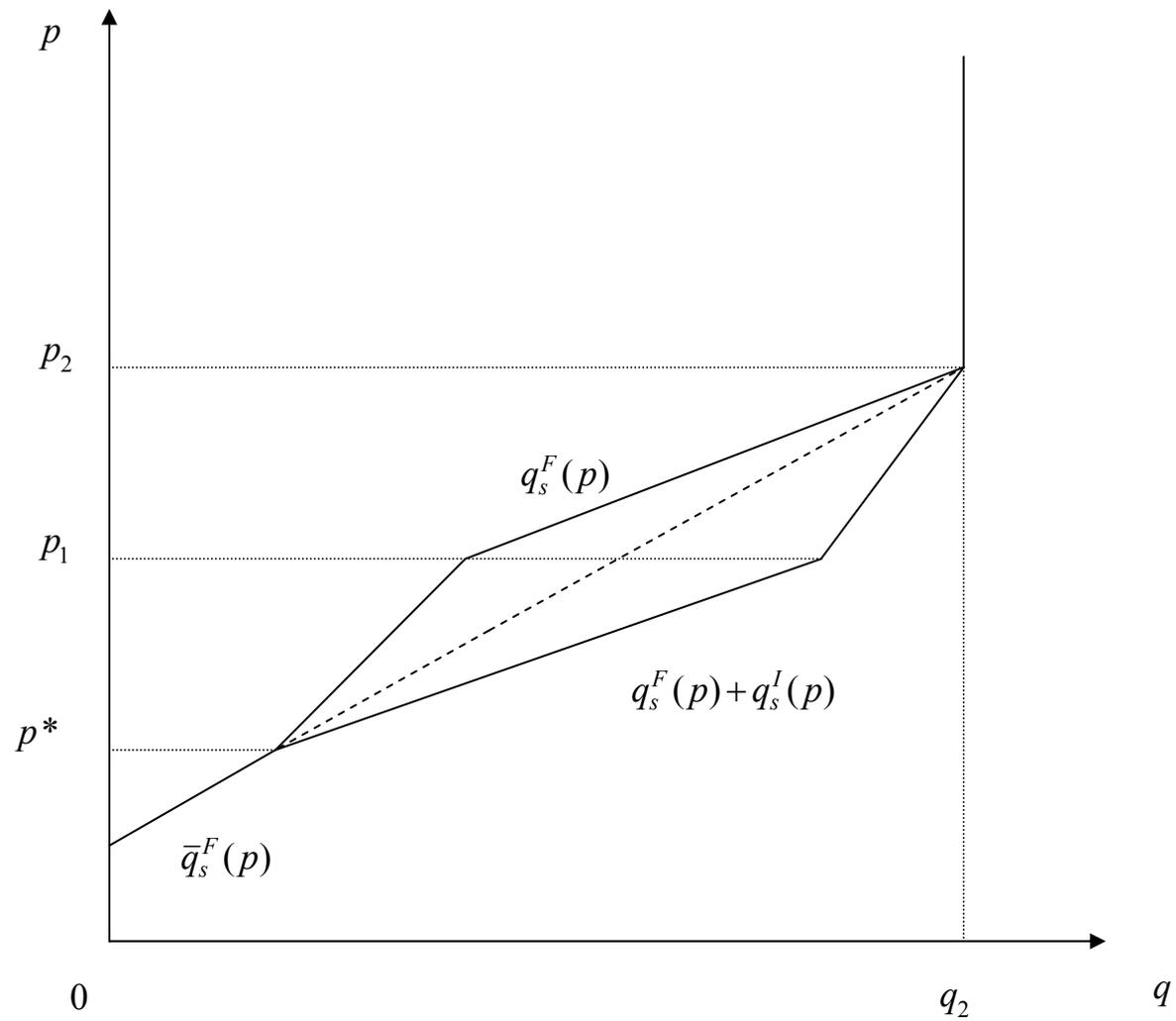


Fig. 2 Formal and Informal Supply with Uniform λ

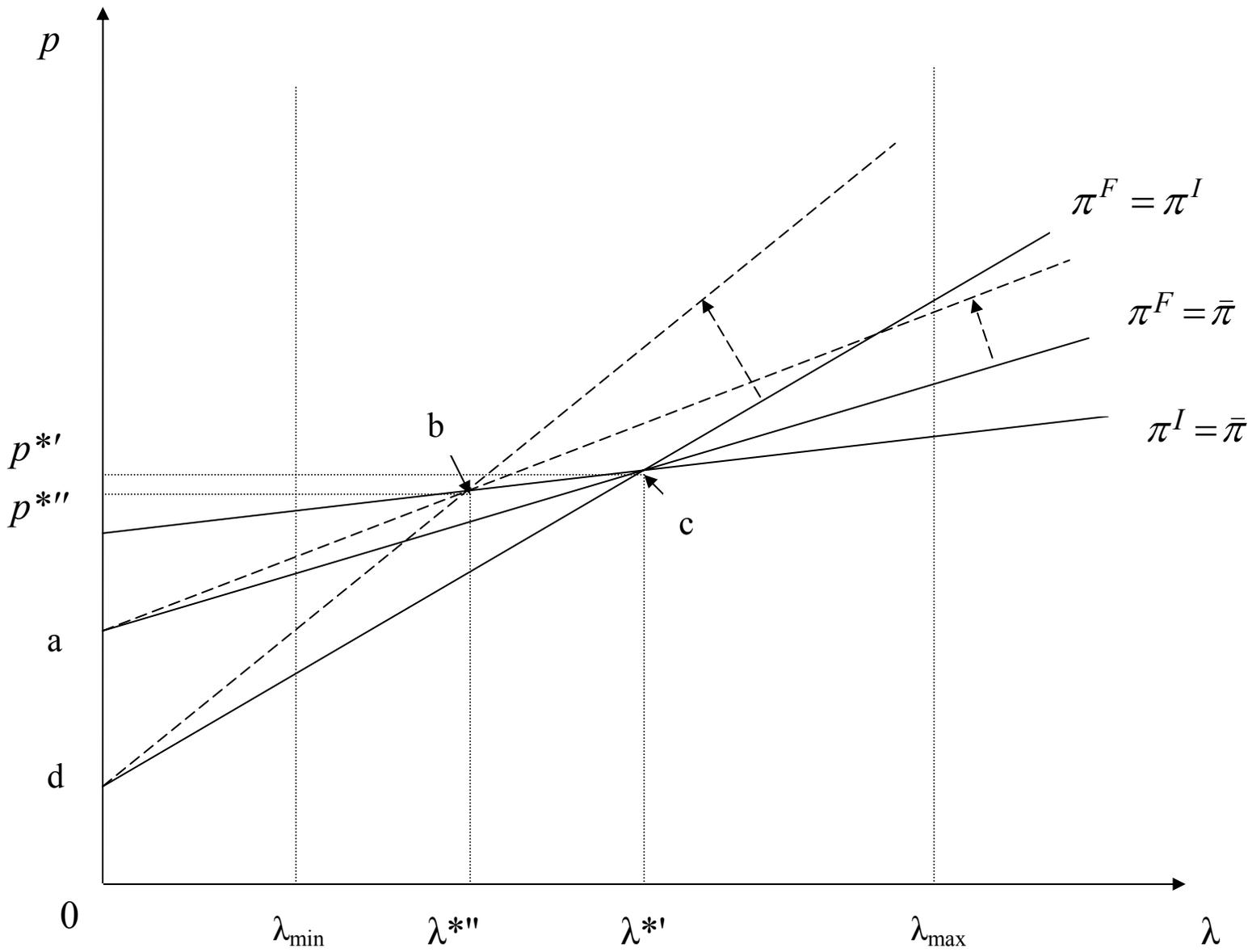


Fig. 3 An Increase in s

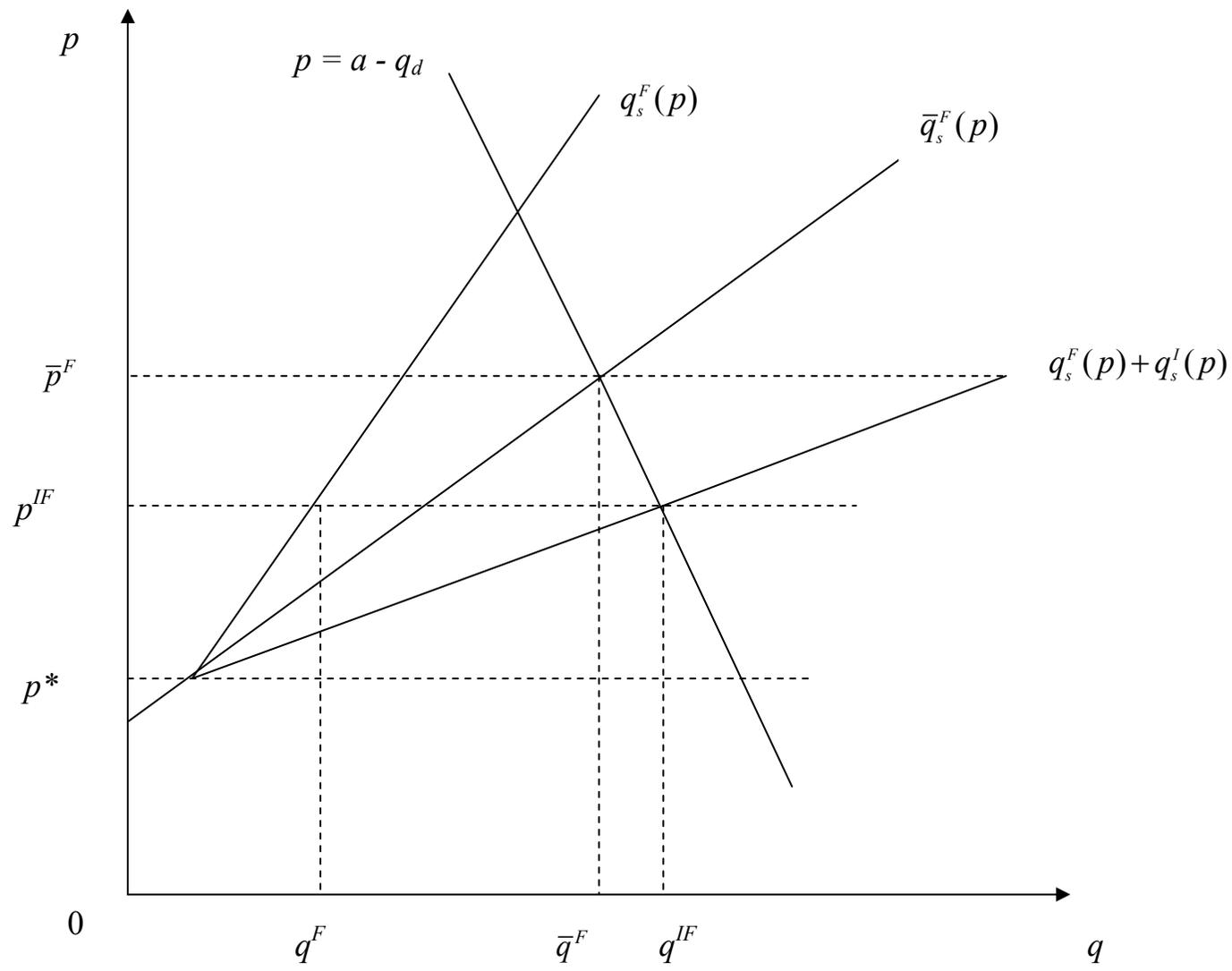


Fig. 4 The Impact of Informality