



15/12
12

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Illicit Firm Behavior: Collusion, Exclusion, and Shadow-Economic Activity

Liliane Karlinger

*Thesis submitted for assessment with a view to obtaining
the degree of Doctor of the European University Institute*

Florence
December 2005

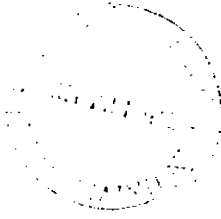
European University Institute



3 0001 0047 4166 8

2
Rosa

LIB
338.8015
1 KAR



EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

ISTITUTO UNIVERSITARIO EUROPEO
08 DIC. 2005
BIBLIOTECA

Illicit Firm Behavior: Collusion, Exclusion, and Shadow-Economic Activity

Liliane Karlinger

The Thesis Committee consists of:

Oriana Bandiera, London School of Economics

Pascal Courty, European University Institute

Massimo Motta, Supervisor, European University Institute

Patrick Rey, University of Toulouse I

Illicit Firm Behavior: Collusion, Exclusion, and Shadow-Economic Activity

Liliane Karlinger
European University Institute

December 2005

This dissertation is dedicated to my Dad, and to Nicola.

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many people. Any attempt to give full credit to each of them is bound to fail, but that does not discourage me from giving it a try. My warmest thanks go to Massimo Motta, for challenging me with so many new and exciting ideas, some of which I finally managed to understand; for encouraging me to pursue my research, when I badly needed encouragement; for never putting me under pressure, when pressure was the last thing I needed. I am also indebted to Patrick Rey, who supervised me during my visiting stay in Toulouse, and who was kind enough to continue offering his support long after I had left Toulouse. I have not yet mentioned those who make sure everything runs smoothly at the Economics Department of EUI, Jessica, Martin, and Manuela, to name a few (those who had to suffer most from my silly questions and never-ending problems), and who made EUI that unforgettable experience to me. Finally, and with all my heart, grazie to Vincenzo, and gracias to Fernando, and thank you to all our friends of Via Sant'Orsola, for opening the doors of their home - and of their hearts - to me, no matter what time of the day or night. And yes, thank you Nicola for telling me about this fabulous place in Florence and encouraging me to apply, back then in Vienna, when I was not even dreaming of ever being accepted to this program, and to Katha and Ingrid, and Nicola again, for remaining faithful friends despite the many years I spent abroad.



Contents

I	Introduction	vi
II	Chapters	1
1	Can Firms Make Themselves Better off by Ignoring Information on Their Competitors?	1
	Introduction	1
	The model	7
	Comparison to the <i>Green-Porter</i> model	13
	The sustainability of collusion	13
	Length of punishment and value of collusion	20
	Welfare Implications	23
	Conclusion	24
	Appendix A: Proofs	29
2	Exclusionary Pricing and Rebates in a Network Industry	32
	Introduction	32
	The setup	36
	Two benchmark price regimes	40
	Uniform linear pricing	40

CONTENTS

iv

Third-degree price discrimination	43
Rebate schemes (Second-degree price discrimination)	51
Welfare Analysis	60
Concluding remarks	64
Appendix A: Proofs	68
Appendix B: The case of full discrimination among buyers	81
Appendix C: Uniform two-part tariffs	86
Miscoordination Equilibria	86
Entry Equilibria	88
Conclusion	100
Welfare Analysis	103
 3 The Underground Economy Across Countries: Evading Taxes, or Evading	
Competition?	105
Introduction	105
The Model	110
Description of the Game	110
The Firms	110
The Tax Authority	113
Solution of the Game	114
Equilibrium in the Product Market (stage 2)	115
Equilibrium at the Entry-Stage of the Game (stage 1)	117

Comparative Statics 119

Discussion 120

Empirical Evidence 125

 Description of the Data 125

 Regression Results 129

 Robustness 132

Conclusion 135

Appendix A: Proofs 141

Appendix B: Descriptive Statistics and First-Stage Regressions 146

Appendix C: Robustness Checks 151

Part I

Introduction

Every day, firms have to make numerous decisions: which price to charge, which quantity and quality to offer, which production technology to use, whether to make an investment or not, and the like.

When making these decisions, the firm's interests may come into conflict with society's objectives. Therefore, numerous laws and regulations try to bring firm behavior in line with social welfare. Cases in point are the labor law, environmental regulation, product and safety standards, and of course competition policy.

Yet, these regulations can resolve the conflict between firms and society only in part: First of all, firms may simply ignore or consciously break the law, in particular if law enforcement is inadequate. Second, in many economic environments regulation is difficult and perforce incomplete. Such twilight zones leave scope for undesirable, yet legal firm behavior.

Illicit firm behavior is a feature of everyday life, and challenges public policy to strike the right balance between economic freedom and needful intervention.

Various strands of economic literature highlight different aspects of the multifaceted problem of illicit firm behavior. A vast literature in industrial organization deals with market power, and what firms do to procure it, either by teaming up with their competitors (in the form of mergers, cartels, or tacit collusion) or by squeezing them out of the market (e.g. through exclusive dealing contracts, predatory pricing, or foreclosure).

Market power is problematic whenever it discourages consumption and/or leads to inefficient production. Apart from such deadweight losses, market power can raise distributive concerns: higher profits for firms have their counterpart in lower consumer welfare.

Chapters 1 and 2 of this thesis want to contribute to this literature. Chapter 1 deals with "tacit collusion", i.e. the silent coordination of price setting among firms. Chapter 2 addresses exclusionary behavior, i.e. a dominant firm engaging in practices which are aimed at deterring entry by a more efficient firm.

Finally, Chapter 3 speaks to the literature on shadow-economic activities, i.e. economic activities which are concealed from public authorities to avoid the payment of taxes and social security contributions, and to avoid compliance with certain legal standards (e.g. labor market regulations, trade licenses).

The shadow economy raises two major concerns: First, undeclared economic activities reduce the tax base, which undermines the financing of public goods and social protection. If the government reacts to the erosion of the tax base by raising tax rates even further, the economy can get trapped in a vicious circle.

Second, firms operating in the underground economy generally do not have access to public contract-enforcement institutions and to credit markets; thus, these firms tend to be confined to inefficient small-scale operation, with adverse consequences for aggregate productivity growth.

The three chapters of this thesis are self-contained and can be read independently. In what follows, I will briefly describe the central research question and the main contribution of each of these chapters.

Chapter 1 analyzes the mechanisms at work when firms try to weaken competition by tacitly coordinating prices in an oligopolistic market. A key assumption in many studies of

tacit collusion is that firms can perfectly monitor each other. However, the fact that there are indeed industries where, because of prevailing business practices, firms are not able to observe their rivals' behavior directly, challenged economic theory to shed light on the crucial role of information in sustaining tacit collusion (Stigler (1964), Green and Porter (1984)).

The conventional wisdom based on this literature is that "more information is always better than less". But then, we may wonder why firms in such markets do not try to improve transparency in one way or another. Obviously, there are many obstacles (legal and incentive-wise) to direct information exchange among firms, but firms might for instance jointly set up an independent information agency at the beginning of the game which has the only purpose of collecting and disseminating information about pricing behavior. Ex ante, every firm should agree to create such an agency, knowing that it will reduce ex-post incentives to deviate, thus helping to sustain the collusive outcome.

However, there are only few practical examples of such "information agencies", and most of them are not even sponsored by the firms. I will claim that there might be good reasons for this apparent lack of such agencies: Even if such an agency was costless and provided reliable (though imperfect) information about competitors' behavior, its presence may not facilitate collusion in any way.

In more technical terms, I analyze the scope for tacit collusion when the outcome itself is not publicly observable, but instead firms directly receive noisy public signals about the actions played. At first sight, it may seem that the better firms can observe each others' pricing behavior, the more likely it is that tacit collusion can be sustained. I find that this

is indeed true if the probability of low-demand states is high.

On the other hand (and this may come as a surprise), if the probability of negative demand shocks is low, there are actually cases where tacit collusion will be more difficult to sustain (in the sense that firms will have to be more patient) than without any information on rivals' behavior. The reason is that in order to take advantage of the information that becomes available, firms need to soften the threat of punishment, which may increase the temptation to undercut the rival, thus creating severe incentive problems.

Hence, the effects of increased observability on the industry under consideration are ambiguous. My central result is therefore that if, for a given discount factor, collusion would be sustainable without signals but not when signals are taken into account, then firms are better off if they ignore the signals and punish whenever one firm has zero profits.

Chapter 2 (co-authored with Massimo Motta) deals with rebates, i.e. discounts applicable where a customer exceeds a specified target for sales in a defined period. Such discounts have been suspected by competition authorities of helping dominant firms to artificially foreclose business opportunities for their competitors (see Gyselen, 2003). However, the economic rationale underlying such practices is not yet well understood.

The purpose of Chapter 2 is to explore the exclusionary potential of rebate arrangements in the presence of asymmetric buyers and network externalities. Our work is closely related to the literature on anticompetitive effects of discriminatory pricing by Innes and Sexton (1993, 1994), and on exclusion through exclusive dealing contracts by Segal and Whinston (2000).

We consider an industry composed of an incumbent firm and an entrant, both supplying a network good, where the entrant has lower marginal cost of production than the incumbent. The good is sold to $m+1$ different buyers, m identical small buyers and 1 large buyer.

For buyers to derive positive utility from consuming a firm's network good, this network must reach a certain minimum size, where a firm's network size is the sum of all sales that this firm makes. We assume that the incumbent disposes of an installed base, and so its network has reached this minimum size already, while the entrant's network has size zero at the outset. In order to reach the minimum size, the entrant has to attract the large buyer plus at least one small buyer.

The central result of this chapter is that rebates may allow the incumbent to break entry equilibria where the incumbent could not have done so under uniform flat prices. For a wide range of parameter values, only the miscoordination equilibrium survives when rebates can be used. Exclusion is more likely to be feasible if the efficiency gap between the two firms is not too wide.

The reason is that the incumbent has an installed base that provides its network with the minimum size, so it can serve all buyers who want to buy from it, no matter how many (or how few) they are, while the entrant can only serve its buyers if it attracts at the least large buyer plus one small buyer. Thus, if the large buyer decides to patronize the incumbent, then the small buyers have no other choice than to buy from the incumbent as well, and vice versa: If the small buyers prefer to buy from the incumbent, then the large buyer will be forced to do so as well, even if he prefers to buy from the entrant. Now, rebates allow

the incumbent to play the two groups of buyers off against each other, which prevents them from coordinating on the more efficient supplier, and so entry will fail.

Chapter 3 addresses the shadow economy, i.e. economic activities which are concealed from public authorities to avoid the payment of taxes and social security contributions, and to avoid compliance with certain legal standards (e.g. labor market regulations, trade licenses). Unreported activities are a universal feature of economic life, and assume considerable proportions even in the industrialized world, where they are estimated to range between 8 and as much as 28 percent of official GDP.

It has been observed that the size of the underground economy (as a fraction of overall economic activity) varies considerably across countries, which has motivated an extensive literature investigating the causes of this particular form of regulation failure: The burden of taxes and social security contributions, extensive labor market regulation, as well as ineffective law enforcement and corruption, have been discussed at length to explain cross-country variations (see Schneider/Enste (2000), Johnson et al (1998), Lemieux et al (1994)).

While acknowledging that all of these factors do play an important role in determining the size of the underground economy, there are good reasons to believe that this list is not exhaustive. In particular, these factors do not explain the substantial variations in the share of the underground economy that have been observed not only across countries, but even within a single country, i.e. within the same legal and institutional framework. A well-documented example is the South of Italy, where the share of the underground economy is twice the national average (De Rita/Camusi (2003)).

In this chapter, I present a novel rationale for the variations in the share of the underground economy which can explain both inter-regional and inter-sectoral differences and sheds new light on cross-country evidence: the intensity of market competition among firms.

The reasoning is as follows: A firm which operates in the underground economy can buy its inputs, in particular labor, at a lower price (by avoiding payroll taxes, not complying with safety and health standards, etc.), thereby reducing its variable cost relative to a firm in the official economy. The underground firm can pass on its savings to consumers, which will reduce market prices, and as a result its competitors' profits fall. Thus, the official firm is put at a competitive disadvantage, and may have to choose between operating underground as well, or going out of business. The keener is competition, the higher is the pressure to reduce costs, and the more likely are underground activities to spread in the industry.

I present cross-country evidence on the impact of entry and competition characteristics on the size of the underground economy in various OECD, transition and developing countries. While the reliability of data on the underground sector is a controversial issue, my regression results indicate that more intense competition is indeed correlated with a higher incidence of shadow-economic activity, thus lending support to my model predictions.

Part II

Chapters

Chapter 1

Can Firms Make Themselves Better off by Ignoring Information on Their Competitors?

Introduction

One of the major achievements of the theory of industrial organization has been to uncover the mechanisms at work when firms try to outflank competition by tacitly colluding in an oligopolistic market. Today, the factors that facilitate such tacit collusion, like high concentration, firm symmetry, high frequency of orders, and multi-market contacts, are well-understood.

A key assumption in most of these analyses is that firms can perfectly monitor each other. However, the fact that there are indeed industries where, because of prevailing business practices, firms are *not* able to observe their rivals' behavior directly, challenged economic theory to shed light on the crucial role of information in sustaining tacit collusion.

In his seminal paper of 1964, Stigler first analyzed the case of a Bertrand-type oligopoly

with stochastic demand where each firm's prices are unobservable to its competitors (i.e. each firm can grant secret price cuts to its customers). Stigler (1964) concluded that without observability of prices, collusion will in general be more difficult to sustain, but can still arise if the cartel provides the right incentives.

Following Stigler's (1964) approach, Green and Porter (1984) developed their model to show that if firms choose quantities (rather than prices) and can only observe the prevailing market price (but not firm-specific or industry supply), then episodes of high industry output (above the collusive level) need not be the result of a collapse of collusion, but should rather be interpreted as part of the firms' equilibrium strategies to ensure tacit collusion in a non-cooperative framework.

Stigler's (1964) and Green-Porter's (1984) work inspired a growing literature on firm behavior under non-observability of competitors' actions. In particular, Abreu, Pearce and Stacchetti (1986) analyzed optimal punishment strategies in oligopolies with imperfect monitoring, showing that every symmetric sequential equilibrium payoff in the *Green-Porter* model can be supported by sequential equilibria having an extremely simple intertemporal structure.

Fudenberg, Levine and Maskin (1994) identify conditions for the folk theorem to apply in repeated games in which players observe a *public* outcome (e.g. the market price in the context of the *Green-Porter* model) that imperfectly signals the (unobservable) actions played (e.g. the quantities set by individual firms).¹

¹Note that it is crucial that the signal be *publicly* observable, because the signal serves two distinct purposes here: on the one hand, it provides information to the agents (which could be achieved by a *private* signal as well), but on the other hand, it also allows firms to *coordinate* their behavior on the signal's

The conventional wisdom based on this literature is that "more information is *always* better than less". But then, we may wonder why firms in such markets do not try to improve transparency in one way or another. Obviously, there are many obstacles (legal and incentive-wise) to direct information exchange among firms, but firms might for instance jointly set up an independent information agency at the beginning of the game which has the only purpose of collecting and disseminating information about pricing behavior. Ex ante, every firm should agree to create such an agency, knowing that it will reduce ex-post incentives to deviate, thus helping to sustain the collusive outcome.

However, as we will see soon, there are only few practical examples of such "information agencies", and most of them are not even sponsored by the firms. I will argue that there might be good reasons for this apparent lack of such agencies. My claim is that even if such an agency was costless and provided reliable (though imperfect) information about competitors' behavior, its presence may not facilitate collusion in any way.

In more technical terms, the purpose of this chapter is to analyze the scope for tacit collusion when the outcome itself is *not* publicly observable, but instead firms *directly* receive public signals about the actions played. This model specification seems more appropriate for the type of industry that we are interested in here, namely a market where firms set *prices* (rather than quantities), and neither prices nor aggregate market demand are publicly observable. Moreover, there is no voluntary information exchange between firms (let alone explicit agreements on behavior), but firms still tend to have at least partial insight into

realizations (which is not the case for a (noisy) private signal, as then the state of the world will no longer be common knowledge among the agents, making it difficult or even impossible to coordinate their actions).

their competitors' behavior.

This information becomes available without any effort on the firm's part. Examples for such "exogenously provided" information include:

- *Cartels*: The idea that cartels would act as "policemen" enforcing collusive outcomes is already highlighted in Stigler (1964). One case in point would be the Joint Executive Committee as analyzed by Porter (1983), which primarily gathered quantity and price information in an attempt to identify deviations by members. However, this evidence dates from a period before the Sherman Act was passed; today, cartels are outlawed in most countries, and hence current examples are rare.
- *Governmental or Consumer Information publications*: Government authorities may decide to publish contract specifications or invoice prices in an attempt to "make the market more transparent". Examples are US railroad grain rates in the 1980's (see Fuller, Ruppel and Bessler (1990), Ruppel and Fuller (1992), and Schmitz and Fuller (1995)) and the Danish ready-mixed concrete market in the early 1990's (see Albæk, Møllgaard and Overgaard (1997)). Most of the time, however, this type of information will be provided by consumer protection agencies which compare price offers for many industries and make this information available to the general public.²
- *New trading technologies*: The internet is thought of having changed the informational structure and trading practices in many markets (see for instance, Klemperer (2000) on internet sales versus dealer sales of cars).

²see Møllgaard and Overgaard (2000), p. 3 f., for examples including telephone services, health and car insurance, pension schemes, home-theater hardware etc.

Note that in the examples given above, the price information becomes *publicly* observable, i.e. each firm in the industry now has some idea about the (previously unobservable) decisions taken by its competitors.

In the following, we will try to understand how such an industry compares to one in which prices are completely unobservable. At first sight, it may seem that the case of partial observability is just an in-between case which probably shares its properties with the two polar cases of perfect observability and complete unobservability.

However, I will show that this is not necessarily the case: in particular, there is no "monotonic" relationship between the degree of observability and the sustainability of tacit collusion. Instead, my results point to the fact that a model of partial observability will generally have very idiosyncratic features that are not trivially implied by the well-understood models of perfect observability and complete unobservability.

The chapter is organized as follows: Section 1 introduces the model; in Section 1, I compare my model to the *Green-Porter* one in terms of sustainability of collusion, length of punishment and value of collusion. Finally, I discuss some policy implications in Section 1 and conclude in Section 1.

Remark

An obvious alternative to having information being supplied exogenously would be to let the firms themselves unilaterally "spy" on each other. The following examples will illustrate this point:

- *Poaching competitors' workers*: in each company, there are key workers who have sensi-

tive information about this company (accountants, controllers, etc.). Competitors can try to make very attractive offers to these workers (in terms of wages, promotion etc.) if these workers will in turn provide the relevant information to their new employer.

- *Corruption*: there are government authorities who have superior information compared to the market participants (e.g. tax authorities, competition authorities); firms can try to bribe them to receive information about their competitors.³
- *Mergers*: a firm can merge with or buy other firms that have relevant information about competitors. Such firms can be e.g. upstream suppliers of competitors or downstream distributors. Alternatively, suppose there is a call for tenders regarding shares of a firm's competitor. Then, the firm might not even be interested in buying these shares, but could still participate in the tendering procedure in order to obtain access to confidential documents which will be distributed to potential buyers during the tender.

Now, one would expect that if information is obtained through unilateral efforts, the insights would be *private* knowledge of the "spying" firm (in particular, if there is room for mistakes or ambiguities in what the "spying" firm can observe, then the firm that was "spied on" would not know what conclusions its rival arrived at). Unfortunately, the properties of repeated games with imperfect monitoring and *privately* observed signals are not yet well understood; one way to resolve the coordination problem in such a model is to allow for

³For an insightful analysis of the relationship between oligopoly and corruption, see Ades and Di Tella (1999)

communication between players, as shown in Kandori and Matsushima (1998) and Compte (1998).

As was mentioned above, this chapter mainly focuses on *tacit* collusion, which excludes explicit communication and coordination between firms. Hence, I will restrict attention to the case of *public* signals (i.e. the first type of information flows as illustrated above) and defer the case of private signals for future research.

The model

In the exposition of the model, I will closely follow Tirole's 1988 treatment of the price-setting variant of *Green-Porter's* model in terms of notation and line of reasoning. I consider an infinitely repeated duopoly game where two symmetric firms, S_i and S_j , produce perfect substitutes at constant marginal cost c . The firms choose prices every period, and *consumers* can perfectly observe these prices so that they will all buy from the low-price firm. The latter assumption may seem somewhat strong, given that *firms* cannot observe each other's prices. However, it will approximately hold in situations where the buyers are large firms searching the market for potential input providers, or private households considering an important purchase (like a car, a home, a holiday trip) and shopping around for the best offer.⁴

Demand for the product is stochastic; with probability p_l , demand will be zero in a

⁴Recently, a new strand of literature evolved, studying oligopolistic markets where consumers can only *imperfectly* observe sellers' prices. One interesting result of this work is that increasing market transparency may not be unambiguously beneficial for consumers (see Nilsson (1999) for a search cost approach, Møllgaard and Overgaard (2000) and (2001), and Schultz (2001) for a product-differentiation approach, and Klempere (2000) for an auction-theoretic approach). Assuming instead *perfect* observability for buyers has its analytical advantages, as it allows us to abstract from these issues on the consumer side of the market and focus solely on firm interaction.

given period ("low-demand state"), and with probability $1 - p_l$, demand will be positive ("high-demand state"). Realizations are assumed to be iid over time.

For the high-demand state, denote the monopoly (or collusive) price by p^m and the per-period monopoly profits by Π^m . I assume that demand is split if the two firms charge the same price. Thus, in a period of high demand, each firm's profit under collusion will be $\Pi^m/2$. Next-period's profits are discounted at rate δ .

If a firm does not sell anything at some date, it does not know *a priori* whether this is due to a low realization of demand or to his competitor charging a lower price. Each firm can however observe its own profits; thus, it is always common knowledge that at least one firm realized zero profits (because then either demand is low, hence the other firm realized zero profits as well, or the other firm undercut).⁵

Moreover (and this is where I depart from *Green-Porter's* model), I assume that after each period, each firm receives a (noisy) signal about his competitor's pricing behavior in that period.⁶ These signals are iid over time and independent of the state of demand, and can be characterized as follows: In a fraction α_i of all cases where S_j behaved collusively (i.e. where $p_j = p^m$), firm S_i understands that its competitor, S_j , did not undercut (i.e. S_i observes $p_j = p^m$ correctly); however, in a fraction of $1 - \alpha_i$ of these cases, S_i receives a wrong signal, indicating that S_j defected (i.e. S_i observes some $p_j < p^m$ when in fact $p_j = p^m$). Analogously, in a fraction β_i of all cases where S_j undercuts, S_i realizes this correctly, whereas in a fraction of $1 - \beta_i$ of these cases, S_i gets it wrong and thinks that low

⁵This common knowledge property will be crucial here, because, as we shall see soon, it allows firms to coordinate their punishment on profit realizations even though the latter are only *privately* observable.

⁶Assume for simplicity that there is no direct cost involved for the firms to obtain these signals.

demand has been realized. By symmetry of the firms, we have $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$. The table below summarizes these properties.

$\frac{1}{2} \leq \alpha_i \leq 1$	S_i 's signal:	S_i 's signal:
$\frac{1}{2} \leq \beta_i \leq 1$	S_j did not defect	S_j defected
S_j did not defect	α_i	$1 - \alpha_i$
S_j defected	$1 - \beta_i$	β_i

I further assume that the realization of S_i 's signal is observable to S_j and vice versa, and that this is common knowledge as well.⁷ Again, this may seem like a strong assumption given that S_i 's prices themselves are not observable for S_j , but it is necessary for the coordination of punishment among firms.⁸ Moreover, if a firm realizes profits $\Pi^m/2$, this implies that it can perfectly infer the other firm's behavior and vice versa. Then, it is common knowledge that demand was high and both firms cooperated, and so they will ignore their signals.

I will now analyze equilibria with the following strategies: There is a collusive phase and a punishment phase. The game starts in the collusive phase. Both firms charge p^m until one firm makes zero profits *and* observes a signal indicating defection⁹ (recall that by the above assumptions, this event will be observed by both firms). The occurrence of this event triggers the punishment phase: both firms will charge c for exactly T periods. If T is finite,

⁷ Hence, these signals are more similar to the publicly provided information described in the Introduction than to "spying" activities on the firm level.

⁸ Recall that, otherwise, we would have to allow for communication (or some other coordination device) between them, which we want to exclude here as we are only interested in the scope for *tacit* collusion.

⁹ This signal may be the firm's own signal or the competitor's signal. Thus, punishment is triggered if *at least one* firm receives a signal indicating defection.

then after T periods both firms will revert to the collusive phase and charge p^m again until the next punishment phase is triggered.

Note that we can allow for alternative definitions of the event that triggers punishment (e.g. "if *exactly one* firm receives a signal..."); in this case the definitions of α (equation (1.1)) and β (equation (1.4)) need to be changed accordingly.¹⁰ Observe also that the *Green-Porter* model corresponds to the case where behavior is *not conditioned* on signals at all, i.e. punishment is *always* triggered whenever zero profits are realized, even in the case where none of the firms receives a signal indicating defection.

Now, let the probability that punishment will be triggered if demand was low and nobody defected be denoted by $1 - \alpha$. If the strategy is such that punishment will be triggered whenever *at least one* firm receives a signal indicating defection, then $1 - \alpha$ corresponds to

$$1 - \alpha = \alpha_i (1 - \alpha_j) + (1 - \alpha_i) \alpha_j + (1 - \alpha_i) (1 - \alpha_j) = 1 - \alpha_i^2$$

and hence

$$\alpha = \alpha_i \alpha_j = \alpha_i^2 \quad (1.1)$$

Thus, the expected discounted present value of the collusive phase is:¹¹

$$V^+ = (1 - p_l) (\Pi^m/2 + \delta V^+) + p_l \alpha (\delta V^+) + p_l (1 - \alpha) (\delta V^-) \quad (1.2)$$

This equation has the following interpretation: under collusion, the high-demand state will be realized with probability $1 - p_l$, each firm will make profits of $\Pi^m/2$ in this period,

¹⁰Note, however, that punishing only when *both* firms get a negative signal will not allow to sustain collusion (just observe that in the case of *perfect* observability, a firm considering unilateral deviation will not face *any* threat of punishment!)

¹¹compare equation (1.2) to the analogous one in the *Green-Porter* model:

$V^+ = (1 - p_l) (\Pi^m/2 + \delta V^+) + p_l \delta V^-$, i.e. there is no chance that the collusive phase continues once low demand has been realized.

and the game will be in the collusive phase next period as well; with probability p_l , the low state of demand will be realized, and both firms make zero profits.

However, in my model the realization of low demand does not automatically trigger the punishment phase; instead, with probability α , each firm will receive the correct signal indicating that the rival did not undercut, and hence the collusive phase will continue. Still, with probability $1 - \alpha$, at least one firm will receive the (wrong) signal that the competitor defected, and consequently the punishment phase starts.

The punishment phase lasts for T periods, after which both firms return to collusive behavior. During the punishment phase, both firms make zero per-period profits. Hence, the expected discounted present value of the punishment phase is:¹²

$$V^- = \delta^T V^+ \quad (1.3)$$

Suppose now that one firm defects while the other continues to charge p^m . Then, if punishment will be triggered with the probability β that at least one firm receives a signal indicating defection, β is defined by:

$$\begin{aligned} \beta &= \beta_i \alpha_i + \beta_i (1 - \alpha_i) + (1 - \beta_i) (1 - \alpha_i) \\ &= 1 - \alpha_i + \alpha_i \beta_i \end{aligned} \quad (1.4)$$

Now, for collusive behavior to be sustainable in this non-cooperative framework, defection must be less profitable than collusion; hence, the following incentive constraint (IC) must

¹²note that this is the same as in the *Green-Porter* model

hold.¹³

$$V^+ \geq (1 - p_l) \Pi^m + (1 - \beta) (\delta V^+) + \beta (\delta V^-) \quad (1.5)$$

The incentive constraint says that the value of collusion (V^+) must be higher than the value of defection. The latter is composed of the one-period profit from undercutting (which is the entire monopoly profit, Π^m , with probability $1 - p_l$, and zero with probability p_l) and the value of next period (which is δV^+ if the game remains in the collusive phase, and δV^- if punishment is triggered). If this constraint holds, no firm will have an incentive to defect (i.e. undercut the competitor during a collusive phase).

Now, we can rewrite equations (1.2) and (1.3) to obtain:

$$V^+ = \frac{(1 - p_l) \Pi^m / 2}{1 - \delta (1 - p_l + p_l \alpha) - \delta^{T+1} (1 - \alpha) p_l} \quad (1.6)$$

and

$$V^- = \frac{\delta^T (1 - p_l) \Pi^m / 2}{1 - \delta (1 - p_l + p_l \alpha) - \delta^{T+1} (1 - \alpha) p_l} \quad (1.7)$$

Rearranging the incentive constraint, (1.5), using equations (1.6) and (1.7), we obtain the incentive constraint in terms of parameters α , β , δ , and p_l :

$$1 + \delta (1 - \beta - 2p_l \alpha) \leq 2\delta (1 - p_l) + \delta^{T+1} [(1 - \alpha) 2p_l - \beta] \quad (1.8)$$

The analogous condition in the *Green-Porter* model is:

$$1 \leq 2\delta (1 - p_l) + \delta^{T+1} (2p_l - 1) \quad (1.9)$$

¹³compare again to the *Green-Porter* model where the incentive constraint reads: $V^- \geq (1 - p_l) \Pi^m + \delta V^-$. Hence, retaliation was certain to occur after any defection, which is no longer the case in our model either.

Hence, our first observation is that the *Green-Porter* model corresponds exactly to my model for parameter values $\alpha = 0$ and $\beta = 1$, for which equations (1.2) and (1.3) as well as the incentive constraint (1.8) will be exactly equivalent to the *Green-Porter* framework.

Now, S_i 's program can be stated as:

$$\max V^+ \text{ s.t. the IC (1.8) holds} \quad (1.10)$$

We see from equation (1.6) that V^+ is decreasing in T , meaning that the longer the punishment phases, the smaller the value for S_i . Hence, S_i 's program is solved by finding the lowest T that just satisfies the IC, condition (1.8), provided such a T exists. This optimal length of punishment will be denoted T^{opt} in the following.

Comparison to the *Green-Porter* model

The sustainability of collusion

The first question we may want to ask is: will tacit collusion be more easy or more difficult to sustain in this model of partial observability compared to the standard *Green-Porter* model?

To answer this question, I will first compare the parameter restrictions that need to hold to guarantee sustainability under the most severe threat possible, i.e. given $T = \infty$. An "educated guess" would probably be that the better firms can observe each others' pricing behavior, the more likely it is that collusion can be sustained. Hence, it may come as a surprise that this is not always the case, as I will show in the following.

For $T = \infty$, the incentive constraint in the *Green-Porter* model reduces to:

$$\delta \geq \frac{1}{2(1-p_l)} \quad (1.11)$$

If this inequality is satisfied, then there exists a $T^{opt} > 0$ (finite or infinite) such that collusion can be sustained. A natural restriction on δ , the discount factor, is: $\delta < 1$. Hence, condition (1.11) can only hold for values of p_l satisfying

$$p_l < \frac{1}{2}$$

Proposition 1 *Increased observability of pricing behavior allows for tacit collusion to be potentially sustainable even for $p_l \geq \frac{1}{2}$, i.e. in cases where collusion would have been impossible in the Green-Porter framework.*

Proof: From the incentive constraint, (1.8), we can derive the counterpart to condition (1.11) in our framework:

$$\delta \geq \frac{1}{2(1-p_l) - 1 + \beta + 2\alpha p_l} \quad (1.12)$$

For this inequality to be consistent with $\delta < 1$, we need to impose

$$p_l < \frac{1}{2} \frac{\beta}{1-\alpha} \quad (1.13)$$

Note that if $\frac{\beta}{1-\alpha} > 1$, then condition (1.13) (which is *necessary*, but not sufficient for sustainability of collusion) is satisfied even for $p_l \geq \frac{1}{2}$. Hence, collusion in our model is potentially sustainable for *any* value of p_l (not just for $p_l < \frac{1}{2}$), provided the associated α is high enough.

□

Note also that for the case of α and β as defined in (1.1) and (1.4), we can never have $\frac{\beta}{1-\alpha} < 1$, as this would imply $\beta_i < 1 - \alpha_i$, which cannot be the case since, by assumption, $\beta_i \geq \frac{1}{2}$ and $\alpha_i \geq \frac{1}{2}$. Thus, all values of p_i that are consistent with collusion in the *Green-Porter* model will *always* be consistent with collusion for our type of strategies as well.

Next, suppose collusion is potentially sustainable in both models. Then, we may wonder if the minimum discount factor δ required to sustain collusion is smaller or higher in our model than in the *Green-Porter* framework. Now, we find that both cases are possible.

Proposition 2 *Even if prices are partially observable, collusion may be more difficult to sustain than in Green-Porter's model of complete unobservability in the sense that the critical discount factor will be higher.*

Proof: If the denominator in (1.12) is smaller than the denominator in (1.11), then the minimum value of δ required to ensure sustainability of collusion is higher in our model than in the *Green-Porter* model, i.e. collusion will be more difficult to sustain. In particular, we have $2(1 - p_i) - 1 + \beta + 2\alpha p_i < 2(1 - p_i)$ whenever

$$p_i < \frac{1}{2} \frac{1 - \beta}{\alpha} \quad (1.14)$$

Note that condition (1.14) is more likely to be satisfied the lower α is.¹⁴ \square

This result qualifies our "educated guess" made above: increased observability does not facilitate tacit collusion in a monotonic way.

¹⁴Note that this result is in some sense an interesting parallel to Overgaard and Møllgaard (2000): in their model, where prices are fully observable by firms but not by consumers, increasing market transparency (i.e. observability of prices by *consumers*) can actually *facilitate* collusion if the number of firms in the market is small (contradicting the conventional consumer protection view that transparency will always be beneficial for consumers); in our model, we see that given fully informed consumers, there is a range of cases where increasing observability of prices by *firms* may in fact make collusion *more difficult* to sustain.

Figure (1.1) illustrates Propositions 1 and 2 graphically for the case where punishment is triggered if *at least one* signal indicating defection is observed (i.e. α and β are defined according to equations (1.1) and (1.4)), and, moreover, we have that $\alpha_i = \beta_i$. The x-axis shows α_i , the y-axis shows p_l . Now, the *white* area represents the region where collusion is

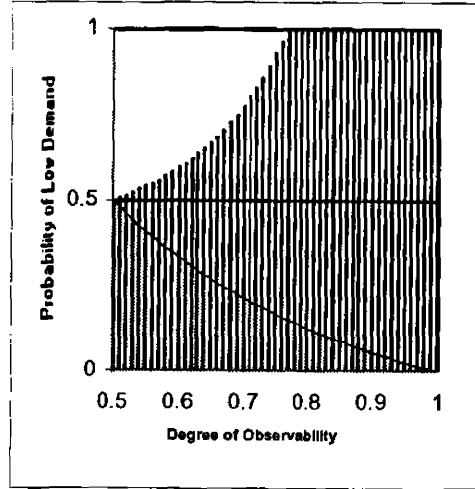


Figure 1.1: Absolute and Relative Sustainability of Collusion

not sustainable at all, while the *hatched* area depicts condition (1.13), i.e. the range of values of p_l that are *consistent* with collusion. Note that for a non-informative signal (i.e. $\alpha_i = \frac{1}{2}$), the relevant interval is $p_l \in [0, \frac{1}{2})$, i.e. coincides with what the corresponding interval in the *Green-Porter* model is, while for any informative signal, this interval will be larger, meaning that the potential for collusion expands as the signal's accuracy improves. The grey triangle in Figure (1.1) illustrates condition (1.14), i.e. the area for which collusion is more difficult to sustain in our model than in *Green-Porter's* model.

Hence, we should expect that whenever, for a given δ , collusion would be sustainable

in *Green-Porter* but not if strategies are conditioned on the signals, then, for reasons of efficiency, firms will prefer to behave à la *Green-Porter* (rather than, say, à la *Bertrand*, which is of course still a feasible equilibrium as well).¹⁵ In such a case, it would be optimal for firms to deliberately ignore signals that become available to them for free.

Lemma 3 *Let collusion be sustainable under Green-Porter strategies, i.e. condition (1.9) is satisfied. Then, the Green-Porter strategies (i.e. start punishment whenever one firm has zero profits, independently of any signal) are still feasible equilibria even if signals are present.*

Proof: see Appendix A

Proposition 4 *If, for a given δ , collusion were sustainable in Green-Porter but not when signals are taken into account, then firms are better off if they punish whenever one firm has zero profits, no matter what the signals indicate about the competitors' behavior.*

Proof: If

$$\frac{1}{2(1-p_l) - 1 + \beta + 2\alpha p_l} > \delta \geq \frac{1}{2(1-p_l)}$$

(which can be the case whenever condition (1.14) is satisfied), then it follows from Proposition 1 that collusion is sustainable in the *Green-Porter* framework but not in ours. Thus, behaving as in *Green-Porter* is not only feasible (as shown in Lemma 3), but also clearly more beneficial for firms. \square

¹⁵In section 1, we will show that if both *Green-Porter* and conditioning on signals are feasible equilibria, then the latter are clearly *superior* in terms of payoffs.

We can therefore conclude that even if signals are rather reliable and strategies sophisticated, firms may not be able to improve on the expected payoffs that *Green-Porter* strategies would already yield. This last result highlights the fact that the *Green-Porter* equilibrium is amazingly efficient, considering how little firms know about each other and how severe their inference problem therefore is.

Discussion

For an economic interpretation of Proposition 2, let us reconsider our incentive constraint (condition (1.5)): When $T = \infty$ (i.e. the length of punishment is infinite), then the value of the punishment phase is zero (see equation (1.3)). Hence, the incentive constraint, (1.5), reduces to:

$$V^+ \Big|_{\alpha \geq \frac{1}{2}} \geq (1 - p_l) \Pi^m + (1 - \beta) (\delta V^+) \quad (1.15)$$

Compared to the corresponding IC in the *Green-Porter* model, i.e.

$$V^+ \Big|_{\alpha = 0} \geq (1 - p_l) \Pi^m$$

we see that conditioning punishment on the signals has two effects on condition (1.15). First of all, observability enters directly on the right-hand side of the inequality through $1 - \beta$, thus adding a strictly positive term to the value of defection (unless $\beta = 1$). If $1 - \beta > 0$, this means that a firm which undercuts has a chance to get away with it (which is not the case in the *Green-Porter* model), and so the temptation to defect increases.

However, the value of collusion, V^+ , is now a function of α (compare with equation

(1.6)). In particular, we have that

$$\frac{\partial V^+}{\partial \alpha} > 0 \text{ for any } T > 0 \quad (1.16)$$

i.e. the value of collusion is in fact increasing in the degree of observability,¹⁶ because the probability that "false alarm" will trigger a price war after a collusive period decreases, thus making it *less* attractive to defect. It depends on the parameter values how these two contrary effects of observability on the incentive constraint will finally play out, and so it is not surprising that there are instances where the net effect is one of increased temptation, thus making it harder or even impossible to sustain collusion.

Inspection of Figure (1.1) as well as condition (1.12) shows that the less informative the signal and the lower the probability of a demand shock, the higher the critical δ must be (relative to the one in *Green-Porter*) to sustain collusion, i.e. firms must be *more* patient to take advantage of the signals. The importance of signal accuracy for sustainability has already been stressed: accuracy increases both the value of collusion and the probability of punishment after defection, thus relaxing the IC.

It seems more surprising that even if the signal is highly accurate, still collusion may be more difficult to sustain if the probability of demand shocks is *low*. We have seen in condition 1.13 that a *high* p_l generally discourages collusion, but relative to *Green-Porter*, a high p_l will actually *facilitate* collusion. The reason is that the benefit of using signals derives precisely from being able to tell apart demand shocks from undercutting. Now, if such demand shocks are very rare, then the losses due to "false alarm" are not too high

¹⁶Note that this result holds in general, i.e. for any given $T > 0$, not just for $T = \infty$.

anyway, and so the IC (1.5) is actually not relaxed by the positive impact of α on V^+ , but instead tightened by the chance (no matter how small) of getting away with defection.

Length of punishment and value of collusion

Suppose now that collusion is sustainable both under our specification and in the *Green-Porter* framework. Then, it would be interesting to know how the outcome in our model compares to the one of *Green-Porter* under the *same* set of parameters $[\delta, p_l]$. In particular, we would like to analyze how the optimal length of punishment evolves as observability improves, and how, given the optimal length of punishment, the value of collusion is affected by partial observability.

Proposition 5 *The optimal length of punishment is decreasing both in α and in β . Whenever collusion is more difficult (more easy) to sustain under partial observability, the corresponding length of punishment will be greater (smaller) than in the Green-Porter model.*

Proof: see Appendix A

We conclude that in our model, price wars will be less frequent than in the *Green-Porter* model, but depending on the level of observability, they may last longer or shorter.

Figure (1.2) exemplifies the result of Proposition 5 for the case of α and β as defined in (1.1) and (1.4), $\alpha_i = \beta_i$ and parameter values $\delta = 0.9$ and $p_l = 0.3$. The thin horizontal line at $T = 3.0887$ represents the optimal length for the *Green-Porter* model, whereas the thick curve shows the optimal lengths for the various values of α_i . Observe that intersection

occurs at $\alpha_i = \frac{1}{1+2p_l} = 0.625$, i.e. at the level of α_i where the minimum δ required to sustain collusion is exactly the same in our model as in the *Green-Porter* model. Now, recall that

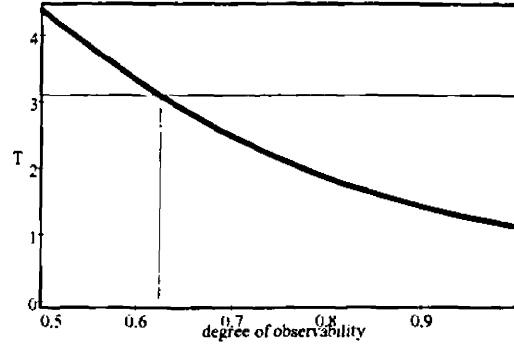


Figure 1.2: Optimal length of punishment

I have already alluded to the gains of increased observability in terms of value of collusion several times (see, e.g., Section 1). So far, however, I only showed that $\frac{\partial V^+}{\partial \alpha} > 0$ for some $T > 0$, see (1.16). Hence, we still need to analyze how the value of collusion as a function of α and β compares to V^+ in the *Green-Porter* model if we take *optimal* punishment into account.

Proposition 6 *Let collusion be sustainable in both models. Provided that the firms use optimal punishment under both settings, the firms will be better off in our model than in the Green-Porter framework whenever $\frac{\beta}{1-\alpha} \geq 1$.*

Proof: First, note that $V^+(T^*)$ in *Green-Porter* is (weakly) smaller than $V^+(T^{**})$ in our model:

$$\frac{(1-p_l)\Pi^m/2}{1-\delta(1-p_l)-\delta^{T^*+1}p_l} \leq \frac{(1-p_l)\Pi^m/2}{1-\delta(1-p_l+p_l\alpha)-\delta^{T^{**}+1}(1-\alpha)p_l}$$

can be rearranged to have

$$\delta^{T^*} \leq \delta^{T^{**}} + \alpha (1 - \delta^{T^{**}})$$

Inserting for T^{**} and T^* and simplifying, we find that the above inequality reduces to

$$\frac{\beta}{1 - \alpha} \geq 1$$

which completes the proof.¹⁷ \square

Figure (1.3) illustrates this fact for the same parameters that were used in Figure (1.2), $\delta = 0.9$ and $p_l = 0.3$, using the values of T^{**} and T^* underlying Figure (1.2) to compute the relevant values of collusion. Again, the thin horizontal line represents V^+ for the *Green-Porter* case, whereas the thick curve shows V^+ for our model as α_i varies. The values on the y-axis are in terms of $\Pi^m/2$. Note that for α and β as defined by equations (1.1) and

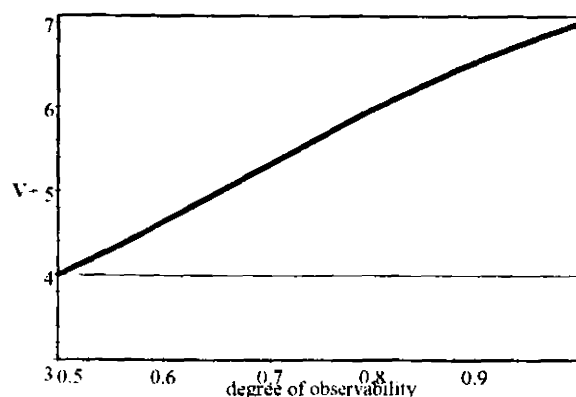


Figure 1.3: Value of Collusion

¹⁷Note that this condition is identical to the one derived in Proposition 1, and so all comments regarding condition (1.13) apply here as well, in particular that condition (1.13) will be satisfied for every reasonable specification of α and β .

(1.4), and $\alpha_i = \beta_i = \frac{1}{2}$, the values of collusion in both models will always be exactly the same, which means that with a non-informative signal, firms can never do better than they do in the *Green-Porter* model of complete unobservability. The difference between the two curves represents the value of the signal for the firms.

Welfare Implications

It is evident that, in terms of total and consumer welfare, the equilibria of the *Green-Porter* model are superior to the models of perfect observability, because collusive monopoly pricing will alternate with periods of marginal-cost pricing. Hence, if firms in an oligopoly cannot observe each others' pricing behavior, this is in fact beneficial for consumers.

Now, how does our model compare to the *Green-Porter* model in terms of consumer welfare? Of course, in the limiting case of $\alpha = \beta = 1$, we are in a situation of perfect observability, and all results developed for this case apply. So let's focus on cases of partial observability, i.e. $\alpha < 1$ and/or $\beta < 1$.

In Section 1, we saw that partial observability allows for collusion to be sustained in cases where it would not have been sustainable under complete unobservability. Thus, partial observability is potentially harmful to consumers when $p_l \geq \frac{1}{2}$.¹⁸

For $p_l < \frac{1}{2}$, I showed that, whenever collusion is sustainable under both types of strategies, then price wars will be less frequent and firms' expected profits will be higher if they condition their behavior on signals (see section (1)). Thus, in the presence of signals, a

¹⁸Note, however, that collusion is less of a problem if p_l is close to 1, because then, consumer demand is low and firms will make zero profits most of the time anyway, while collusion is an important issue if p_l is low, meaning that consumer demand and potential welfare losses are high.

new equilibrium may emerge which dominates both the *Green-Porter* and the Bertrand equilibrium in terms of expected profits, and hence is detrimental to consumer and total welfare.

Moreover, even if collusion is not sustainable when signals are taken into account, firms may simply ignore the signals, thus attaining the same equilibrium that would have been feasible in the absence of signals.

To summarize, even though signals will not necessarily facilitate collusion, it is hard to imagine a situation where these signals could lead to a collapse of collusion which would otherwise have arisen.¹⁹ Hence, to be on the safe side, the competition policy authority should keep observability at the lowest possible level. Then, collusion will be most difficult to sustain (and least profitable) for any value of p_l . One way to fight collusion is of course to encourage entry into the industry, as this will create severe incentive problems for all firms.²⁰

Conclusion

I analyzed a symmetric Bertrand duopoly model with uncertain demand, where one firm's prices are unobservable to its competitors, but firms receive (noisy) public signals about their competitor's pricing behavior. At first sight, it may seem that the better firms can observe each others' pricing behavior, the more likely it is that tacit collusion can be

¹⁹The only difficulty that the firms might face (in particular if signals become available which allow for a "new" equilibrium to be sustained) is how to *coordinate* on the payoff-dominant equilibrium (in particular if that requires switching from a "historical" to a "new" equilibrium).

²⁰This is of course true for tacit collusion in general, with or without signals. However, the benefits of the signals considered in this paper decrease dramatically as the number of firms increases; just observe that generalizing the definition of α (compare to (1.1)) will yield $\alpha(n) = \alpha_i^n$, which rapidly converges to zero as n increases (unless $\alpha_i = 1$), implying that false alarm triggering a price war becomes more and more likely the more firms are in the market.

sustained. Comparing the equilibrium of our model to the benchmark model of *Green-Porter*, where pricing behavior is assumed to be completely unobservable, I first found indeed that increased observability allows for tacit collusion to be sustainable even if the probability of low-demand states is high, i.e. in cases where collusion would have been impossible in the *Green-Porter* framework.

On the other hand (and this may come as a surprise), if the probability of negative demand shocks is low, there are actually cases where tacit collusion will be more difficult to sustain (in the sense that firms will have to be more patient) than in *Green-Porter's* model. The reason is that in order to take advantage of the information that becomes available, firms need to soften the threat of punishment, which may increase the temptation to undercut the rival, thus creating severe incentive problems.

Hence, the effects of increased observability on the industry under consideration are ambiguous. My central result is therefore that if, for a given δ , collusion would be sustainable in *Green-Porter* but not when signals are taken into account, then firms are better off if they ignore the signals and punish whenever one firm has zero profits.

If tacit collusion is sustainable in both models, I found that the optimal length of punishment (i.e. the length of "price wars") is decreasing as observability increases, and will be higher (lower) than in the *Green-Porter* framework if collusion is more difficult (easy) to sustain than in *Green-Porter's* model. In terms of expected profits, the equilibrium under signaling always yields higher payoffs for the firms. Hence, from a welfare-analytical point of view, I conclude that the competition authority should keep observability at the lowest

possible level to make collusion as difficult to sustain as possible.

As far as empirical support for our model is concerned, it is hard to imagine how to derive testable implications and take them to the data, because some of the crucial variables in our model are unobservables. If at all, the model may lend itself to experimental analysis.

On the theoretical level, note that our framework was designed to analyze the main question as directly as possible, and hence it lacks some desirable features that would allow to address a broader scope of issues. Possible extensions of the model include:

- analyzing the case where signals are private rather than public; This model specification would more appropriately describe what I mentioned in the Remark of the Introduction as "spying on each other";
- allowing for such information acquisition to be costly for the firms, where this cost may increase with the signal's degree of precision, in order to obtain results about the optimal level of "market research" that firms will carry out (i.e. endogenize the level of observability) and about the impact of entry on these research efforts;
- allow for asymmetry of the quality of signals between firms, to see if one firm's signal acquisition activity creates positive/negative externalities on the other firm;
- allow for a larger set of equilibrium strategies, varying in the way firms condition their behavior on the signals.

Bibliography

- [1] ABREU, DILIP/PEARCE, DAVID/STACCHETTI, ENNIO (1986): "Optimal Cartel Equilibria with Imperfect Monitoring", *Journal of Economic Theory* 39(1), 251-69.
- [2] ADES, ALBERTO/DI TELLA, RAFAEL (1999): "Rents, Competition, and Corruption", *American Economic Review* 89(4), 982-993.
- [3] ALBÆK, SVEND/MØLLGAARD, H. PETER/OVERGAARD, PER B. (1997): "Government-Assisted Oligopoly Coordination? A Concrete Case", *The Journal of Industrial Economics* 45(4), 429-443.
- [4] COMPTE, OLIVIER (1998): "Communication in Repeated Games with Imperfect Private Monitoring", *Econometrica* 66(3), 597-626.
- [5] FUDENBERG, DREW/LEVINE, DAVID/MASKIN, ERIC (1994): "The Folk Theorem with Imperfect Public Information", *Econometrica* 62(5), 997-1039.
- [6] FULLER, STEPHEN W./RUPPEL, FRED J./BESSLER, DAVID A. (1990): "Effect of Contract Disclosure on Price: Railroad Grain Contracting in the Plains", *Western Journal of Agricultural Economics* 15(2); 265-271.
- [7] GREEN, EDWARD J./PORTER, ROBERT H. (1984): "Noncooperative Collusion under Imperfect Price Information", *Econometrica* 52(1), 87-100.

- [8] KANDORI, MICHIOHRO/MATSUSHIMA, HITOSHI (1998): "Private Observation, Communication and Collusion", *Econometrica* 66(3), 627-652.
- [9] KLEMPERER, PAUL (2000): "Why Every Economist Should Learn Some Auction Theory", *mimeo*, 1-18.
- [10] MØLLGAARD, H. PETER/OVERGAARD, PER B. (2000): "Market Transparency: A Mixed Blessing?", *mimeo*, 1-32.
- [11] MØLLGAARD, H. PETER/OVERGAARD, PER B. (2001): "Market Transparency and Competition Policy", *mimeo*, pp. 1-36.
- [12] NILSSON, ARVID (1999): "Transparency and Competition", *mimeo*, 1-22.
- [13] PORTER, ROBERT H. (1983): "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886", *Bell Journal of Economics* 14(2), 301-314.
- [14] RUPPEL, FRED J./FULLER, STEPHEN W. (1992): "Information Impacts and Determinants of Information Selection: An Experimental Approach", *Journal of Agricultural and Resource Economics* 17(1), 150-161.
- [15] SCHMITZ, JOHN/FULLER, STEPHEN W. (1995): "Effect of Contract Disclosure on Railroad Grain Rates: An Analysis of Corn Belt Corridors", *Logistics and Transportation Review* 31(2), 97-124.
- [16] SCHULTZ, CHRISTIAN (2001): "Transparency and Tacit Collusion", *mimeo*, 1-27.
- [17] STIGLER, GEORGE J. (1964): "A Theory of Oligopoly", *The Journal of Political Economy* 72(1), 44-61.
- [18] TIROLE, JEAN (1988): "The theory of industrial organization", Cambridge : MIT Press.

Appendix A: Proofs

Proof of Proposition 5: Part 1: To check how T^{opt} behaves as α and β vary, let's analyze condition (1.8). Recall from the firm's maximization problem, (1.10), that the smallest T that satisfies (1.8) is the optimal length of punishment. Hence, let (1.8) hold with equality so that

$$1 + \delta (1 - \beta - 2p_l \alpha) - 2\delta (1 - p_l) - \delta^{T+1} [(1 - \alpha) 2p_l - \beta] = 0$$

defines T^{opt} as a function of α and β . Then, after solving for T^{opt} explicitly and differentiating with respect to α and β , we obtain (using $p_l < \frac{1}{2}$ and (1.13), which insure sustainability of collusion in both models)

$$\frac{\partial T^{opt}}{\partial \beta} = \frac{\delta^{T+1} - \delta}{\ln(\delta) \delta^{T+1} [(1 - \alpha) 2p_l - \beta]} < 0$$

and

$$\frac{\partial T^{opt}}{\partial \alpha} = 2p_l \frac{\partial T^{opt}}{\partial \beta} < 0$$

We see that the higher the probability that punishment will be avoided if no defection occurred (i.e. the higher α), and the higher the probability that punishment will indeed be triggered if defection occurred (i.e. the higher β), the shorter the punishment phase will be.

Part 2: Denote by T^* the optimal length of punishment in the *Green-Porter* model (i.e. the T for which condition (1.9) holds with equality), and by T^{**} the corresponding variable in our model. Then, we find that $T^{**} \geq T^*$ implies that

$$p_l \leq \frac{1}{2} \frac{1 - \beta}{\alpha}$$

which corresponds exactly to condition (1.14). Recall that whenever condition (1.14) holds, collusion will be more difficult to sustain in a model of partial observability than in *Green-Porter's* model. \square

Proof of Lemma 3: Suppose firm j follows the *Green-Porter* strategies even though signals are available. Then, we need to check if firm i has an incentive to deviate by conditioning its behavior on the signals instead.

Now, suppose firm i devised a strategy that differs from j 's, and let us analyze the type of deviations that might be profitable for i . Denote the deviating firm's value of collusion by V_i^+ , its value of defection by V_i^D , and its value of punishment by V_i^- . Then, we can characterize the possible deviations by the following three cases²¹:

Case (i): Would firm i ever want to continue cooperation in a situation where j 's strategy requires punishment (e.g. if zero profits were realized but both signals indicate that no defection occurred)? The answer is no: if firm j sets $p_j = c$, then it is rational for firm i to set $p_i = c$ as well (even though any $p_i > c$ would of course be a best reply as well).

Case (ii): Would firm i ever want to punish in a situation where j 's strategy requires continuation of (or return to) cooperation? Again, the answer is no: any $\tilde{p}_i \in (c, p^m)$ would yield a strictly higher payoff than $p_i = c$, i.e.

$$(1 - p_l)\Pi(\tilde{p}_i) + \delta V_i^- > 0 + \delta V_i^-$$

Moreover, we can conclude from cases (i) and (ii) that $V_i^- = \delta^{T^*} V_i^+$, where T^* is the number

²¹Note that by the One-Period-Deviation Criterion (cf. *Tirolé*, 1988, p. 265), it is *sufficient* to show that deviating *once* is not profitable to conclude that *no* finite or infinite sequence of deviations can ever be profitable.

of periods that firm j will punish (i.e. the optimal length of punishment for the *Green-Porter* model).

Case (iii): Would firm i ever want to deviate otherwise in a situation where j 's strategy requires continuation of cooperation? Clearly, firm i 's optimal deviation is to undercut firm j slightly, thus realizing a deviation payoff of $V_i^D = (1 - p_l)\Pi^m + \delta V_i^-$. Firm i 's incentive constraint therefore reads:

$$V_i^+ \geq (1 - p_l)\Pi^m + \delta V_i^-$$

Inserting for V_i^- and rearranging yields $V_i^+ \geq \frac{1}{1 - \delta^{T^*+1}}(1 - p_l)\Pi^m$. Now, if $V_i^+ < V^+$, i.e. firm i 's strategy yields a strictly lower payoff than the *Green-Porter* type strategy followed by firm j (and still available to firm i), then firm i 's strategy cannot be a profitable deviation (since $V_i^+ < V^+$ implies $V_i^- < V^-$ and $V_i^D < V^D$). Hence, we must have $V_i^+ \geq V^+$. But then, notice that we have

$$V_i^+ \geq V^+ \geq \frac{1}{1 - \delta^{T^*+1}}(1 - p_l)\Pi^m$$

where the second inequality now represents the incentive constraint for *Green-Porter* type strategies, which holds by assumption. But then, we must also have that firm i 's IC is satisfied, since the right-hand side of the ICs is the same for both types of strategies.

To conclude, firm i will behave exactly like firm j , i.e. there is no circumstance under which firm i would want to deviate. Hence, by symmetry, the same must be true for firm j , and so ignoring the signals and behaving à la *Green-Porter* is still an equilibrium even if signals are available. \square

Chapter 2

Exclusionary Pricing and Rebates in a Network Industry

Introduction

Rebates, i.e. discounts applicable where a customer exceeds a specified target for sales in a defined period, have been suspected by competition authorities of helping dominant firms to artificially foreclose business opportunities for their competitors (see Gyselen, 2003). However, the economic rationale underlying such practices is not yet well understood.

The purpose of this chapter is to explore the exclusionary potential of rebate arrangements in the presence of network externalities. We consider an industry composed of an incumbent firm and an entrant, both supplying a network good, where the entrant has lower marginal cost of production than the incumbent.

The good is sold to $m + 1$ different buyers, m identical small buyers and 1 large buyer. Buyers' valuation for the good is increasing in the number of other buyers buying that same good as well (i.e. in the size of the network). The two networks are not compatible with each other, i.e. network externalities can only arise among customers of the same firm.

For buyers to derive positive utility from consuming a firm's network good, this network

must reach a certain minimum size. We assume that the incumbent disposes of an installed base, and so its network has reached this minimum size already, while the entrant's network has size zero at the outset. In order to reach the minimum size, the entrant has to attract the large buyer plus at least one small buyer.

The timing of the game is as follows: First, both the incumbent and the entrant simultaneously announce their (binding) offers; once these offers have become common knowledge, each buyer decides which firm to patronize, and how much to buy from this firm.

We consider three price regimes: (i) uniform flat prices, (ii) third-degree price discrimination under two-part tariffs, and (iii) rebate schemes (i.e. second-degree price discrimination under two-part tariffs). Under the rebate scheme, firms can only discriminate among buyers by the quantity they buy, but not by their size or identity.

If firms can only use uniform flat prices, then the game has two equilibria: entry equilibria, where the entrant undercuts the incumbent, and all buyers buy from the entrant; and miscoordination equilibria, where all firms buy from the incumbent, although the entrant makes a better offer to them. Either of the two equilibria can arise under all parameter values. Under third-degree price discrimination with two-part tariffs, the miscoordination equilibria continue to exist for all parameter values, while the entry equilibrium will only exist if the entrant is sufficiently more efficient than the incumbent.

The situation is similar if firms cannot openly discriminate among buyers, i.e. if they are restricted to rebate schemes. The central result of this chapter is that rebates may allow the incumbent to break entry equilibria where the incumbent could not have done so

under uniform flat prices. For a wide range of parameter values, only the miscoordination equilibrium survives when rebates can be used. Exclusion is more likely to be feasible if the efficiency gap between the two firms is not too wide.

The reason is that rebate schemes, when applied to buyers who differ in size, will be a tool of (de facto) discrimination, even if the schemes as such are uniform. Now, the possibility of discriminating between buyers and redistributing rents between them should help the entrant challenge the incumbent just as much as it helps the incumbent defend its monopoly position.

However, the incumbent has an installed base that provides its network with the minimum size, so it can serve all buyers who want to buy from it, no matter how many (or how few) they are, while the entrant can only serve its buyers if it attracts at the least large buyer plus one small buyer. Thus, if the large buyer decides to patronize the incumbent, then the small buyers have no other choice than to buy from the incumbent as well, and vice versa: If the small buyers prefer to buy from the incumbent, then the large buyer will be forced to do so as well, even if he prefers to buy from the entrant. Now, rebates allow the incumbent to play the two groups of buyers off against each other, which prevents them from coordinating on the more efficient supplier, and so entry will fail.

We also find that there is a trade-off between maximizing the entrant's chances to enter, and minimizing welfare losses (no matter which firm eventually serves the buyers). Discriminatory two-part tariffs raise the highest barriers to entry, but yield the most efficient outcome (full efficiency if the entrant serves, lowest-possible inefficiency if the incumbent serves). The

opposite is true for linear tariffs (lowest entry barriers, but least efficient outcomes).

Uniform rebates are somewhere between these two extremes. Note, however, that they are sufficient to achieve full efficiency. In other words, if the entrant is sufficiently efficient to enter under uniform rebates, then allowing discriminatory two-part tariffs will not yield any efficiency gains, but may jeopardize entry.

Our work is closely related to Innes and Sexton (1993, 1994), who also analyze the anticompetitive potential of discriminatory pricing. In their papers, however, there is uncertainty about entry (in particular about the cost of entry and the potential efficiency gains) at the time when buyers and incumbent interact. Moreover, the contractual instruments available to the agents are very sophisticated, whereas in our study, the strategic variables are as parsimonious as possible. Finally, buyers are identical in Innes and Sexton, so that there is no role for buyer asymmetry as we analyze it in our framework.

Our work is also related to Segal and Whinston (2000), who show the exclusionary potential of exclusive contracts when the incumbent can discriminate on the compensatory offers it makes to buyers. Our study differs from theirs in several respects: (i) in their game the incumbent has a strategic advantage in that it is allowed to contract with buyers before entry occurs; (ii) if buyers accept the exclusivity offer of the incumbent, they commit to it and cannot renegotiate it even if entry occurs; (iii) buyers are symmetric and only linear pricing is considered. In our paper, instead, (i) the incumbent and the entrant choose price schedules simultaneously, (ii) buyers simply observe prices and decide on whom to buy, without having to commit to buying from one or the other; (iii) we explore the role of rebates

and quantity discounts in a world where buyers of differing sizes exist. Yet, the mechanisms which lead to exclusion in the two papers are very similar (both papers present issues of buyers' miscoordination, and scale economies which are created by fixed costs in their model are created instead by network effects in ours).

The setup

Consider an industry composed of two firms, the incumbent I , and an entrant E . The incumbent supplies a network good, and has an installed consumer base of size $\beta_I > 0$. I incurs constant marginal cost $c_I \in (0, \frac{1}{2})$ for each unit it produces of the network good.

The entrant can supply a competing network good at marginal cost $c_E = 0$, so that $c_E < c_I$, i.e. the entrant is more efficient than the incumbent in supplying the good. E has not been active in the market so far, i.e. it has installed base $\beta_E = 0$, but it can start supplying the good any time; in particular, there is no need to sink any fixed costs of entry first.

The good can be sold to $m + 1$ different buyers, indexed by $j = 1, \dots, m + 1$. There are m identical small buyers, and 1 large buyer. Goods acquired by one buyer cannot be resold to another buyer, but they can be disposed of at no cost by the buyer who bought them (in case the latter cannot consume them). Side payments of any kind between buyers are ruled out. Define firm i 's network size s_i (where $i = I, E$) as

$$s_i = \beta_i + q_i^1 + \dots + q_i^{m+1}$$

i.e. the firm's installed base plus its total sales to all "new" buyers.

The large buyer's demand for firm i 's network good at unit price $p_i^l \geq 0$ is given by

$$q_i^l(p_i^l) = \begin{cases} \max\{(1-K)(1-p_i^l), 0\} & \text{if } s_i \geq \bar{s} \\ 0 & \text{if } s_i < \bar{s} \end{cases} \quad (2.1)$$

while a typical small buyer's demand for firm i 's network good at unit price $p_i^s \geq 0$ is

$$q_i^s(p_i^s) = \begin{cases} \max\{\frac{K}{m}(1-p_i^s), 0\} & \text{if } s_i \geq \bar{s} \\ 0 & \text{if } s_i < \bar{s} \end{cases} \quad (2.2)$$

The parameter $K \in (0, 1)$ is an indicator of the relative weight of the small buyers in total market size: if all buyers buy at the same unit price (i.e. if $p_i^l = p_i^s$), then $1 - K$ measures the large buyer's market share, while K measures the market share of the group of small buyers. Assume that $1 - K > \frac{K}{m}$, so that if offered the same price, the large buyer's demand is always larger than a typical small buyer's demand.¹

Note that the assumption $1 - K > \frac{K}{m}$ implies an upper bound on K , namely

$$K < \frac{m}{m+1} \in \left[\frac{1}{2}, 1\right)$$

and that total potential market size is fixed at 1,

$$m\frac{K}{m} + (1 - K) = 1.$$

Our demand functions are identical across buyers up to the size factor ($1 - K$ or $\frac{K}{m}$), so that a monopolist who could charge discriminatory linear prices would set a uniform unit price $p_i^* = \frac{1}{2}(1 + c_i)$. Recall that $c_E = 0$, so that $p_E^* = \frac{1}{2}$; then, our assumption that $c_I < \frac{1}{2}$ implies that the entrant is never radically more efficient than the incumbent.

¹Later, we will allow firm i 's unit prices to differ across buyers of different size.

If firm i 's network size s_i is below the threshold level \bar{s} , no buyer (neither large nor small) would want to buy firm i 's good. We assume that

$$\beta_I \geq \bar{s}$$

i.e. the incumbent has already reached the minimum size, while the entrant's installed base is $\beta_E = 0$. In order to operate successfully, the entrant will have to attract enough buyers to reach \bar{s} .²

Key Assumption: *In order to reach the minimum size, the entrant has to serve the large buyer plus at least one small buyer:*

$$\bar{s} > \max \{1 - K, K\} \quad (2.3)$$

Thus, winning the large buyer's orders is indispensable for the entrant to operate successfully. However, neither demand of the large buyer alone, nor demand of all small buyers taken together, is sufficient for the entrant to reach the minimum size.

Note that only units which are actually consumed by a buyer count towards firm i 's network size. The demand functions also define the quantities that buyers can at most consume, namely 0 if $s_i < \bar{s}$, and $1 - K$ (for the large buyer) or K/m (for the small buyer) if $s_i \geq \bar{s}$. We do not allow E to produce units and throw them away (or give them away for free to buyers who cannot consume them), in order to reach the minimum size.

We also assume that the threshold level \bar{s} is weakly below the total potential market size. Thus, if the entrant gets to sell to all $m + 1$ buyers at marginal cost, then it will reach

²Note that if the entrant manages to reach the minimum size \bar{s} , then consumers will consider I 's and E 's networks as being of homogenous quality, even if $s_I \neq s_E$.

the minimum size for sure:

$$\bar{s} \leq mq_E^s(c_E) + q_E^l(c_E) = 1 \quad (2.4)$$

Note that inequality (2.4) together with $c_E < c_I$ imply that the social planner would want the entrant (and not the incumbent) to serve all buyers.

Play occurs in the following sequence:

$t = 0$: The incumbent and the entrant simultaneously announce their price schemes, which will be binding in $t = 1$.³

$t = 1$: Each of the $m + 1$ buyers decides whether to patronize the incumbent or the entrant.

As for the price schemes that firms can offer in $t = 0$, we will consider three possibilities:

- (1) the benchmark case of uniform linear prices (Section 2);
- (2) the second benchmark case of third-degree price discrimination under two-part tariffs (Section 2); and

(2) the case of central interest, that is second-degree price discrimination under two-part tariffs, i.e. uniform quantity discounts or "rebates" (Section 2).

(Appendix B briefly discusses the case of full price discrimination. A possible third benchmark is considered in Appendix C, namely uniform two-part tariffs, where firms can offer fixed payments (ruled out in Case (1)), but cannot discriminate among buyers, neither by their type (Case (2)) nor by the quantity they buy (Case (3)).)

³In other words, firms can *commit* to make fixed payments at the end of $t = 1$ (buyers don't have to be concerned that firms renege on payments), so we exclude the possibility that firms just make vague promises, keeping buyers in the dark about how much they will actually get in the end (though such situations may arise in practice, see Gyselen, 2003).

Let us assume that offers are observable to everyone, e.g. because they have to be posted publicly. Then, when the buyers have to decide which firm to buy from, the firms' offers will be common knowledge. In $t = 1$, buyers decide which firm to buy from. We will restrict attention to the case where a buyer can only buy from one of the two firms, but not from both of them simultaneously.

Two benchmark price regimes

Uniform linear pricing

As a first benchmark case, let us consider the situation where firms can only use uniform flat prices (but no fixed payments or unit prices which vary with quantity). In line with the work of Bernheim and Whinston (1998) and Segal and Whinston (2000), we find that our game has two types of pure-strategy Nash equilibria: one where all buyers (or sufficiently many) buy from the entrant, and one where all buyers buy from the incumbent.

The following proposition illustrates the simplest of these two types of equilibria.

Proposition 7 (*equilibria under uniform flat prices*) *If firms can only use uniform flat prices, the following two pure-strategy Nash equilibria exist under the continuation equilibria as specified (after eliminating all equilibria where firms play weakly dominated strategies):*

(i) Entry equilibrium:

- if $\bar{s} \leq 1 - c_I$, E sets $p_E = c_I$, I sets $p_I = c_I$, and all buyers, after observing $p_E \leq \min\{p_I, 1 - \bar{s}\}$, buy from E .

- if $\bar{s} > 1 - c_I$, E sets $p_E = 1 - \bar{s}$, I sets $p_I = p_I^*$ (where p_I^* is firm I 's monopoly price).

and all buyers, after observing $p_E \leq \min \{p_I, 1 - \bar{s}\}$, buy from E .

(ii) *Miscoordination equilibrium*: I sets $p_I = p_I^*$, E sets $p_E = p_E^*$ (where p_E^* is firm E 's monopoly price), and all buyers, after observing $p_I - p_E \leq p_I^*$, end up buying from I .

Proof: see Appendix A

Which type of equilibrium will eventually be played depends on the underlying continuation equilibria, i.e. on how buyers coordinate their purchasing decisions after observing the firms' offers⁴: If a buyer can rely on all other buyers patronizing E whenever E 's offer is at least as good as I 's, then it is perfectly rational for this buyer to buy from E as well. This, in turn, corresponds exactly to what all other buyers expected him to do, and so confirms the rationality of their own supplier choice. Under such a continuation equilibrium, the entry equilibrium of Proposition 7 (i) will arise.

If instead each buyer suspects all other buyers to patronize I even when I 's price is strictly higher than E 's (as in the miscoordination equilibrium of Proposition 7 (ii)), then no buyer will want to buy from E : Recall that no individual buyer's demand is ever sufficient for E 's network to reach the minimum size \bar{s} . Then, being the only buyer to buy from E means ending up with a good that has zero value to that buyer (no matter how cheap it is). Hence, as long as buying from I still gives positive surplus, each buyer will want to buy from I , which then confirms all other buyers in their decision to buy from I as well.

In some sense, a buyer who buys from I is always on the safe side: The incumbent's network benefits from its installed base, so that I 's good will always generate strictly positive

⁴where "coordination" describes the collective behavior under *individual* decision making; we do not allow buyers to meet in $t = 1$ and make a joint decision on which firm to patronize.

utility to whoever buys it, no matter what the other buyers do. Under such a continuation equilibrium, the incumbent can even charge its monopoly price (and will optimally do so) without losing the buyers to the entrant.⁵ These equilibria are particularly troublesome, because they show that a highly inefficient market outcome can persist even in the presence of an efficient competitor.

The equilibria characterized in Proposition 7 represent extreme cases, in the sense that the underlying continuation equilibria are the most favorable ones for the firm that serves the buyers in equilibrium. *These equilibria are by no means the only equilibria that can arise in our game.*

For instance, there are other equilibria where all buyers do miscoordinate on the incumbent, but the latter can at most charge some price $\hat{p}_I < p_I^*$. Such an equilibrium can be sustained by continuation equilibria where buyers buy from I as long as $p_I - p_E \leq \hat{p}_I$, but would switch to E if the price difference exceeded \hat{p}_I . Likewise, there are entry equilibria where the entrant must charge a strictly lower price than c_I (or $1 - \bar{s}$) to induce buyers to coordinate on E . *For the rest of the chapter, we will focus on those continuation equilibria which are the most profitable ones for the firm that eventually serves the buyers.*

Finally, there can also be equilibria where both I and E offer the exact same price, and a critical number of buyers patronize E (so that E reaches the minimum size), while the remaining buyers buy from I . These equilibria can only be sustained by very specific continuation equilibria, and we will not consider them in the following sections of this chapter.

⁵In this situation, the entrant is indifferent among all prices $p_E \geq 0$ it could charge, and might as well offer its monopoly price, which weakly dominates all other possible equilibrium prices.

Third-degree price discrimination

As a second benchmark case, suppose that a supplier $i = I, E$ can offer contracts of the type

$$T_i^j = p_i^j q_i - R_i^j, \text{ with } j = s, l$$

where the fixed component R_i^j could be either positive or negative (if $R_i^j < 0$, it is a franchise fee, i.e. a payment from the type j buyer to the firm i ; if $R_i^j > 0$, it is a slotting allowance, i.e. a payment from the firm i to the buyer j), and where p_i^j is the variable component of the tariff.

We assume that the suppliers can discriminate between one group and the other of buyers, but not within each of them (Appendix B briefly discusses this case), and contrary to the case of rebates analyzed in Section 2, there is no possibility of personal arbitrage by buyers: a large buyer cannot 'pretend' to be a small one and vice versa.

Buyers seek to maximize total surplus, which is the sum of net consumer surplus and possible lump-sum payments they receive from or have to pay to the firms. Define net consumer surplus as follows:

$$CS_i^j(p_i^j) = \frac{1}{2} (1 - p_i^j) q_i^j(p_i^j) \quad (2.5)$$

$$= \begin{cases} \frac{1}{2} (1 - K) (1 - p_i^j)^2 & \text{if } s_i \geq \bar{s} \text{ and } p_i^j \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = l \quad (2.6)$$

$$= \begin{cases} \frac{1}{2} \frac{K}{m} (1 - p_i^j)^2 & \text{if } s_i \geq \bar{s} \text{ and } p_i^j \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = s \quad (2.7)$$

Lemma 8 (*miscoordination under third-degree price discrimination*) *For all parameter values, there is an equilibrium where I sets*

$$p_I^s = p_I^l = c_I; R_I^s = -\frac{1}{2} \frac{K}{m} (1 - c_I)^2, R_I^l = -\frac{1}{2} (1 - K) (1 - c_I)^2$$

E makes the analogous offer, $p_E^s = p_E^l = c_E = 0$; $R_E^s = -\frac{1}{2} \frac{K}{m}$, $R_E^l = -\frac{1}{2} (1 - K)$, and all buyers, after observing that I offers non-negative total surplus, buy from I.

Proof: Buyer j is indifferent between buying from I , and not buying at all: buying from I yields total surplus $CS_I^j(c_I) + R_I^j = 0$, and not buying at all yields zero surplus as well. Buying from E , given that all other buyers buy from I (so that E would not reach the minimum size), would yield $0 + R_E^j < 0$, so buyers strictly prefer to buy from I . Given that buyers buy from I as long as $CS_I^j(p_I^j) + R_I^j \geq 0$, the incumbent will optimally set price equal to marginal cost to generate maximum consumer surplus, and then use the fixed component of the tariff to fully extract this surplus, thus making maximal profits $CS_I^l(c_I) + mCS_I^s(c_I)$.

Now, suppose the entrant deviates by offering a strictly positive payment to the small buyers: $p_E^s = p_E^l = c_E = 0$ (wlog); $R_E^s = \varepsilon > 0$, and $R_E^l \leq -mR_E^s < 0$. (The analogous reasoning applies to an offer where instead the large buyer would receive $R_E^l > 0$ while $R_E^s \leq -\frac{1}{m}R_E^l < 0$.) Under this offer, the small buyers will no longer buy from I , as they will rather accept $\frac{K}{m}$ units for free from the entrant to qualify for payment R_E^s . Then, even if the large buyer continues to buy from I , so that the entrant will not reach the minimum size (and hence the $\frac{K}{m}$ units remain without value to the small buyers), they will obtain a strictly positive surplus: $CS_E^s(c_E \mid s_E < \bar{s}) + R_E^s = 0 + \varepsilon > 0$.

However, they will be unable to consume the units they received from E , and so they will dispose of their units: Recall that $q_E^j(p_E^j) = 0$ if $s_E < \bar{s}$, and that $K/m < \bar{s}$, so that no buyer will be able to consume the good if he is the only one to do so. Hence, E 's network still has size zero even after E gave $\frac{K}{m}$ units to each of the small buyers, because only units which are actually consumed count towards a firm's network size. But then, the large buyer will not want to switch to E , as he would obtain strictly negative surplus from E , $CS_E^l(c_E \mid s_E < \bar{s}) + R_E^l \leq 0 - m\varepsilon < 0$.

Thus, the large buyer will continue to buy from I , confirming the small buyers' expectation that E 's network will remain below the minimum size (so that disposing of their units is the only option left to them). But then, E will not break even: if E sells at marginal cost to the small buyers, E cannot make strictly positive payments to them, unless E makes positive profits on the large buyer.

We can conclude that if buyers are miscoordinated, any feasible alternative offer E can make must satisfy $R_E^s \leq 0$ and $R_E^l \leq 0$. But such an offer will not induce the small buyers to leave I unless the large buyer does so as well, which will never happen precisely because buyers are miscoordinated. But then, E might as well offer $p_E^s = p_E^l = c_E = 0$; $R_E^s = -\frac{1}{2}\frac{K}{m}$, $R_E^l = -\frac{1}{2}(1 - K)$, which completes our proof. \square

Now, we want to find the conditions under which there is an entry equilibrium of the game where firm I and E make simultaneous offers T_i^j .

Suppose there is a candidate equilibrium at which the entrant charges:

$$p_E^s = p_E^l = c_E = 0; R_E^s, R_E^l,$$

where the entrant optimally chooses to set the variable price equal to its marginal cost so as to maximize the rents that arise from the relationships with the buyers (otherwise, surplus would be inefficiently lost).

Since the entrant needs both the large buyer and at least one small buyer, this candidate equilibrium would not survive if the incumbent could make an offer that makes either the large buyer or the small buyers better off, so that it becomes a dominant strategy for this type of buyer to buy from I , no matter what the other buyers do. Let us look at each possibility in turn.

The first question is whether the incumbent can profitably induce the large buyer to switch. In this case, the small buyers would be forced to buy from I as well, even if I 's offer to them is much less attractive than E 's offer. The best offer I can make to the large buyer is to extract all the surplus from the small buyers and offer it to the large buyer. Such a deviation would take the form:

$$p_I^s = p_I^l = c_I; R_I^s = -CS_I^s(c_I), R_I^l = mCS_I^s(c_I).$$

Therefore, the candidate equilibrium can survive this deviation only if the entrant leaves the large buyer with a larger payoff than the one offered by the incumbent, that is only if:

$$CS_E^l(c_E) + R_E^l \geq CS_I^l(c_I) + mCS_I^s(c_I),$$

which can be rewritten as:

$$-R_E^l \leq CS_E^l(c_E) - CS_I^l(c_I) - mCS_I^s(c_I). \quad (\text{cond 1})$$

Second, the incumbent may also induce a deviation of the small buyers. To this end, it could extract all the surplus from the large buyer and offer it to the small buyers. Such a deviation would consist of the offer:

$$p_I^s = p_I^l = c_I; R_I^l = -CS_I^l(c_I), R_I^s = CS_I^l(c_I)/m.$$

In order for the candidate equilibrium to survive this deviation, the entrant must therefore make an offer such that:

$$CS_E^s(c_E) + R_E^s \geq CS_I^s(c_I) + CS_I^l(c_I)/m,$$

or:

$$-R_E^s \leq CS_E^s(c_E) - CS_I^s(c_I) - \frac{CS_I^l(c_I)}{m}. \quad (\text{cond 2})$$

Note that conditions (cond 1) and (cond 2) must hold simultaneously. Also note that the entrant's profits must be non-negative: Since, at the candidate equilibrium, the entrant is selling at marginal cost, the following break-even condition must hold:

$$-R_E^l - mR_E^s \geq 0. \quad (\text{cond 3})$$

An entry equilibrium where $p_E^s = p_E^l = c_E = 0$; R_E^s, R_E^l can therefore survive only if conditions (cond 1), (cond 2) and (cond 3) will simultaneously hold. Optimality requires firm E to charge the highest possible fee (or to leave the lowest possible allowance) to the buyers. Therefore, at the optimum conditions (cond 1) and (cond 2) will be binding. By writing (cond 1) and (cond 2) with equality and inserting them in (cond 3) we obtain:

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq 2(CS_I^l(c_I) + mCS_I^s(c_I)).$$

In words, entry equilibria can only arise if the total rent generated by the entrant is at least twice as high as the total rent generated by the incumbent. We can make use of the demand functions of the buyers and insert the actual consumer surpluses into the previous inequality, which can then be simplified to

$$c_I \geq 1 - \sqrt{\frac{1}{2}}.$$

In other words, an entry equilibrium can exist only if the entrant is sufficiently more efficient than the incumbent.

This contrasts sharply with the case of uniform linear tariffs of Section 2, where entry equilibria could arise even if the efficiency gap between the entrant and the incumbent was very small (i.e. even if $c_I \rightarrow c_E = 0$). But discriminatory two-part tariffs allow the incumbent to strategically redistribute rent across different types of buyers in order to exclude the entrant from the market.

Finally, the incumbent's offer in such an entry equilibrium (where the incumbent does not sell anything) will depend on how the buyers coordinate on the entrant: If, for instance, buyers buy from E whenever E 's offer is at least as good as I 's (analogously to the continuation equilibrium of Proposition 7 (i)), then the incumbent's equilibrium offer must exactly match the entrant's, e.g. as follows:

$$p_I^s = p_I^l = c_I; R_I^l = mCS_I^s(c_I), R_I^s = CS_I^l(c_I)/m$$

Note that this offer is not actually feasible (if I sold at marginal cost, it could not afford to make strictly positive payments to all the buyers). But if I offered less than that, the

entrant would want to follow suit and reduce its offers as well, thus violating conditions (cond 1) and/or (cond 2), and such offers could not be sustained as an equilibrium.

Proposition 9 (*entry under third-degree price discrimination*) *If firms can use two-part tariffs and discriminate between large and small buyers (but not among small buyers), then the entry equilibrium can only arise if*

$$c_I \geq 1 - \sqrt{\frac{1}{2}} \simeq 0.2929$$

and is characterized by

$$\begin{aligned} p_E^s &= p_E^l = c_E = 0; R_E^s = -\frac{K - (1 - c_I)^2}{2m}; R_E^l = -\frac{(1 - K) - (1 - c_I)^2}{2} \\ p_I^s &= p_I^l = c_I; R_I^l = K \frac{(1 - c_I)^2}{2}; R_I^s = \frac{1 - K(1 - c_I)^2}{m} \end{aligned}$$

with all buyers buying from the entrant after observing that the entrant's offer is at least as good as the incumbent's.

Proof: follows from above, where the expressions for R_E^s , R_E^l , R_I^s , and R_I^l were obtained by inserting from definition (2.5). \square

It is worth studying whether the entrant offers a positive or a negative fixed component to the buyers at equilibrium. It is straightforward to note that:

$$R_E^s \leq 0 \quad \text{iff} \quad K \geq (1 - c_I)^2; \quad \text{and} \quad R_E^l \leq 0 \quad \text{iff} \quad K \leq 1 - (1 - c_I)^2.$$

We can now illustrate the result in the plane $((1 - c_I)^2, K)$, as in Figure 2.1. The

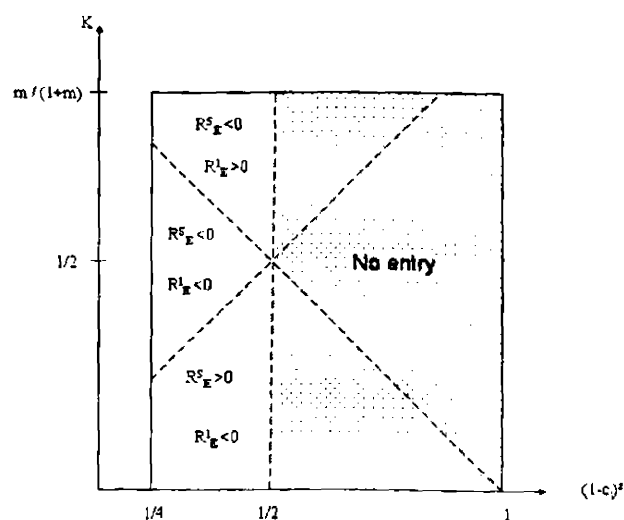


Figure 2.1: Region where the entry equilibrium exists (plain) and does not exist (dotted)

Figure shows the regions where the entry equilibrium exists and characterizes it by showing whether at the equilibrium firm E has to pay or not a fixed fee to the buyers. First of all, note that for any given admissible level of K the more efficient is firm E relative to firm I (that is, as we move horizontally to the left of the plane) the more likely that we find an equilibrium in which firm E is able to extract surplus from both the large and the small buyers.

Let us now look at the comparative statics on K . For any given value of $(1 - c_I)^2$, an increase in K makes it more likely that firm E charges a positive fee to the small buyers and a negative fee to the large buyer. In particular, note that:

$$mR_E^s \leq R_E^l \quad \text{iff} \quad K \geq \frac{1}{2}.$$

In other words, when $K \geq \frac{1}{2}$, that is when the small buyers account for most of the market, the entrant will extract more surplus from them than from the large buyer at equilibrium.

To understand these results, note that according to our assumptions the entrant needs to have purchases from both the small and the large buyer. When K is small, what the incumbent will want to do is to extract as much as possible from the large buyer, whose surplus is larger than the aggregate surplus of the small buyers, to induce the small buyers to buy from it; when K is large, the opposite will occur: the share of the small buyers is the largest, and the incumbent will try to extract as much as possible from them to offer it to the large buyer. Hence, when K is small, the entrant will need to make its best offer to the small buyers, whereas when K is large, it is the small buyers' market share which is largest, and therefore it is the large buyer who needs to be induced to buy away from the incumbent.

Rebate schemes (Second-degree price discrimination)

Let us now consider the case where firms cannot make their offers directly depending on the type of buyer (large or small), but have to make uniform offers to both types which may only depend on the quantity bought by buyer $j = 1, \dots, m+1$:

$$T_i(q_i^j) = \begin{cases} p_{i,1}q_i^j - R_{i,1} & \text{if } q_i^j \leq \bar{q}_{i,1} \\ p_{i,2}q_i^j - R_{i,2} & \text{if } q_i^j \geq \bar{q}_{i,2} \end{cases}$$

The fixed component, $R_{i,1}$ or $R_{i,2}$, can again be either positive or negative.⁶ The difference is that each buyer can now choose his tariff from this price menu by buying either below the sales target $\bar{q}_{i,1}$ or above the sales target $\bar{q}_{i,2}$, where $\bar{q}_{i,1} \leq \bar{q}_{i,2}$.

It is well-known that such quantity discounts or rebates, when applied to buyers who differ in size, will be a tool of (de facto) discrimination, even if the schemes as such are uniform. But to achieve discrimination, the tariffs have to be set in a way that induces buyers to self-select into the right category, with small buyers voluntarily buying the low target, and the large buyer choosing to buy the high target.

First, consider the large buyer $j = l$, and suppose his demand at price $p_{i,2}$ is above the threshold, i.e. $q_i^l(p_{i,2}) \geq \bar{q}_{i,2}$ (this will be the only relevant case). Then, the large buyer can either buy $q_i^l(p_{i,2})$, which yields total surplus $CS_i^l(p_{i,2}) + R_{i,2}$, or he can buy below the threshold $\bar{q}_{i,1}$, i.e. $q_i^l = \min \{q_i^l(p_{i,1}), \bar{q}_{i,1}\}$ at price $p_{i,1}$, in which case his net consumer surplus can be expressed as

$$CS_i^{l,net}(p_{i,1}, \bar{q}_{i,1}) = \begin{cases} CS_i^l(p_{i,1}) & \text{if } s_i \geq \bar{s} \text{ and } q_i^l(p_{i,1}) < \bar{q}_{i,1} \\ \bar{q}_{i,1} \left(1 - p_{i,1} - \bar{q}_{i,1} \frac{1}{2} \frac{1}{1-K}\right) & \text{if } s_i \geq \bar{s} \text{ and } q_i^l(p_{i,1}) \geq \bar{q}_{i,1} \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

Next, consider a typical small buyer $j = s$, and suppose his demand at price $p_{i,2}$ is below the threshold, i.e. $q_i^s(p_{i,2}) < \bar{q}_{i,2}$ (again, this will be the only relevant case). Then, a small buyer may either buy $q_i^s(p_{i,1})$, which yields total surplus $CS_i^s(p_{i,1}) + R_{i,1}$, or he can buy the sales target $\bar{q}_{i,2}$ at price $p_{i,2}$ (i.e. a quantity which exceeds his actual demand at this price).

⁶Results are qualitatively similar if we restrict fixed payments to go from firms to buyers only, while ruling out franchise fees paid by buyers to firms.

If $\bar{q}_{i,2} > \frac{K}{m}$, i.e. if the sales target is above the largest quantity he can consume, $q_i^s(p_i = 0) = \frac{K}{m}$, then the excess units, $\bar{q}_{i,2} - \frac{K}{m}$, can be disposed of at no cost.⁷ Define the small buyer's net consumer surplus of buying $\bar{q}_{i,2}$ units as

$$CS^{s,net}(p_{i,2}, \bar{q}_{i,2}) = \begin{cases} \bar{q}_{i,2} (1 - p_{i,2} - \bar{q}_{i,2} \frac{1}{2} \frac{m}{K}) & \text{if } s_i \geq \bar{s} \text{ and } \bar{q}_{i,2} < \frac{K}{m} \\ \frac{1}{2} \frac{K}{m} - p_{i,2} \bar{q}_{i,2} & \text{if } s_i \geq \bar{s} \text{ and } \bar{q}_{i,2} \geq \frac{K}{m} \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

We say that firm i 's offer satisfies the "self-selection condition" if the large buyer prefers to buy above the threshold, and the small buyers prefer to buy below the threshold, i.e. if

$$CS_i^l(p_{i,2}) + R_{i,2} \geq CS_i^{l,net}(p_{i,1}, \bar{q}_{i,1}) + R_{i,1} \quad (2.10)$$

$$\text{and } CS_i^s(p_{i,1}) + R_{i,1} \geq CS_i^{s,net}(p_{i,2}, \bar{q}_{i,2}) + R_{i,2}$$

For any offer that satisfies the self-selection condition, denote $(p_{i,1}, R_{i,1}, \bar{q}_{i,1})$ by $(p_i^s, R_i^s, \bar{q}_i^s)$, and $(p_{i,2}, R_{i,2}, \bar{q}_{i,2})$ by $(p_i^l, R_i^l, \bar{q}_i^l)$, for $i = I, E$.

Lemma 10 (*miscoordination under rebates*) *For all parameter values, there is an equilibrium where I sets*

$$p_I^s(c_I) = \frac{c_I + \frac{1}{m} (1 - \frac{1}{1-K} \frac{K}{m})}{1 + \frac{1}{m} (1 - \frac{1}{1-K} \frac{K}{m})} > c_I, p_I^l = c_I, \bar{q}_I = q_I^s(p_I^s(c_I))$$

and fully extracts consumer surplus from the small buyers, while leaving some rent to the large buyer:

$$R_I^s = -CS_I^s(p_I^s), R_I^l = -CS_I^l(c_I) + (CS_I^{l,net}(p_I^s, q_I^s(p_I^s)) + R_I^s) < 0$$

⁷ Recall that we excluded reselling of units between buyers (while allowing for free disposal), so the only thing a small buyer can do with units he cannot consume is to throw them away.

*E makes the analogous offer with $0 < p_E^s(c_E) < p_I^s(c_I)$, $p_E^l = c_E = 0$, and all buyers, after observing that *I* offers non-negative total surplus, buy from *I*.*

Proof: see Appendix A

This is the most profitable miscoordination equilibrium under rebates: The incumbent will receive positive fixed payments from both types of buyers (recall that we explicitly allowed for fixed payments to go from buyers to firms), and on top of that, *I* will earn a positive mark-up on the sales to the small buyers. Again, there are alternative miscoordination equilibria (under different continuation equilibria), where the incumbent would make lower profits, and may even have to make positive payments to one or both types of buyers.

Note that the incumbent's profits will be lower under rebates than under third-degree price discrimination: *I* sells above marginal cost to the small buyers, which reduces their consumer surplus (and hence their franchise fee), and *I* does not extract the full consumer surplus $CS_I^l(c_I)$ from the large buyer. Both features of *I*'s equilibrium offer follow directly from the introduction of personal arbitrage.

Suppose that *I* wanted to replicate the more profitable offer under third-degree price discrimination, setting a uniform unit price of $p_{I,1} = p_{I,2} = c_I$, and franchise fees

$$R_{I,1} = -CS_I^s(c_I) \quad \text{if } q_I^j \leq \bar{q}_I = q_I^s(c_I)$$

$$R_{I,2} = -CS_I^l(c_I) \quad \text{if } q_I^j > \bar{q}_I = q_I^s(c_I)$$

This "first-best" offer does not satisfy the self-selection condition, because the large buyer would then prefer the small-buyer tariff, i.e. he would buy $\bar{q}_I = q_I^s(c_I)$ and enjoy strictly positive net surplus $CS^{l,net}(c_I, q_I^s(c_I)) - CS_I^s(c_I) > 0$ (although this means the large buyer is

quantity-constrained, because $q_I^s(c_I) < q_I^l(c_I)$). Thus, arbitrage implies that I 's offer to the large buyer must leave the latter with at least as much rent as buying below the threshold would yield.

This requirement also explains why the incumbent does not simply charge $p_I^s = p_I^l = c_I$ to all buyers to extract the full consumer surplus from the small buyers. The intuition is as follows. Suppose that I charges c_I to both large and small buyers. It can then extract $CS_I^s(c_I)$ from the small buyers, and must leave at least $CS_I^{l,net}(c_I, q_I^s(c_I)) - CS_I^s(c_I)$ to the large buyer.

Now, suppose instead that I raises the price p_I^s by an ε , while leaving $p_I^l = c_I$ unchanged: of course, this means efficiency losses, i.e. the additional profits I makes on sales to the m small buyers are lower than the losses in rent that I can extract from them through the fixed fee. On the other hand, I can now extract more rent from the large buyer: under the higher p_I^s (and accordingly lower sales target), it is much less attractive for the large buyer to buy below the threshold.

Thus, no matter how large K or m , there will always be an ε such that the losses of raising p_I^s (on the small buyers) are more than outweighed by the gains (on the large buyer), and so $p_I^s = c_I$ cannot be optimal.

Next, we want to find the conditions under which there is an entry equilibrium of the game where firm I and E make simultaneous offers $T_i(q_i^j)$.

Suppose there is a candidate equilibrium at which the entrant charges:

$$p_E^s = p_E^l = c_E = 0; R_E^s, R_E^l,$$

where the entrant optimally chooses to set the variable price equal to its marginal cost so as to maximize the rents that arise from the relationships with the buyers (otherwise, surplus would be inefficiently lost). If this offer is to sustain an entry equilibrium, we must have that:

(i) the incumbent cannot make an offer that makes either the large buyer or the small buyers better off, so that it becomes a dominant strategy for this type of buyer to buy from I (as in Section 2); and

(ii) the entrant's offer satisfies the self-selection condition.

ad (i): Recall that the best offer I can make to the large buyer under third-degree price discrimination is to extract all the surplus from the small buyers and offer it to the large buyer:

$$p_I^s = p_I^l = c_I; R_I^s = -CS_I^s(c_I), R_I^l = mCS_I^s(c_I).$$

If this offer satisfies the self-selection condition, then it is also the best offer that I can make under the rebate scheme. But note that personal arbitrage may now be a problem: In particular, the small buyers may want to mimic the large buyer and buy $q_I^l(c_I)$, as long as the extra expenditure on those units above their demand is more than compensated by the payment of R_I^l . In this case, the incumbent will have to make an alternative offer that satisfies the self-selection condition, and such an offer will necessarily generate less than the full surplus, $CS_I^l(c_I) + mCS_I^s(c_I)$, for the large buyer.

Likewise, the incumbent's best offer to the small buyers under third-degree price discrimination, where I extracts all the surplus from the large buyer and offers it to the small

buyers:

$$p_I^s = p_I^l = c_I; R_I^l = -CS_I^l(c_I), R_I^s = CS_I^s(c_I)/m$$

is no longer feasible if arbitrage is possible. The large buyer will always want to buy below the threshold and receive a strictly positive payment R_I^s , rather than buying above the threshold and being left with zero rent. Thus, any incentive-compatible offer that redistributes rents from the large buyer to the small buyers will provide less total surplus than $CS_I^s(c_I) + CS_I^l(c_I)/m$ to each small buyer.

Again, the entrant's offers under an entry equilibrium must satisfy the following necessary conditions:

$$CS_E^s(c_E) + R_E^s \geq CS_I^s(p_I^{s,best}) + R_I^{s,best}$$

$$CS_E^l(c_E) + R_E^l \geq CS_I^l(p_I^{l,best}) + R_I^{l,best}$$

$$-R_E^l - mR_E^s \geq 0$$

where $(p_I^{s,best}, R_I^{s,best})$ denotes I 's best rebate offer to the small buyers (i.e. the offer that maximizes small buyers' total surplus while satisfying the self-selection condition), and analogously, $(p_I^{l,best}, R_I^{l,best})$ denotes I 's best rebate offer to the large buyer (see Appendix A for the values which $p_I^{l,best}$, $p_I^{s,best}$, $R_I^{l,best}$, and $R_I^{s,best}$ will take). Inserting into the break-even constraint, we now obtain

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq CS_I^l(p_I^{l,best}) + R_I^{l,best} + m[CS_I^s(p_I^{s,best}) + R_I^{s,best}] \quad (2.11)$$

This condition is weaker than the corresponding condition under third-degree price discrim-

ination:

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq 2 (CS_I^l(c_I) + mCS_I^s(c_I))$$

(We prove in Appendix A that $CS_I^l(p_I^{l,best}) + R_I^{l,best} \leq CS_I^l(c_I) + mCS_I^s(c_I)$ and $CS_I^s(p_I^{s,best}) + R_I^{s,best} < CS_I^s(c_I) + \frac{1}{m}CS_I^l(c_I)$.) In other words, the entrant is less likely to be excluded under rebates than under third-degree price discrimination.

ad (ii): Consider the candidate equilibrium offer

$$p_E^s = p_E^l = c_E = 0$$

$$R_E^s = CS_I^s(p_I^{s,best}) + R_I^{s,best} - CS_E^s(c_E)$$

$$R_E^l = CS_I^l(p_I^{l,best}) + R_I^{l,best} - CS_E^l(c_E)$$

If this offer satisfies the self-selection condition, then the feasibility condition (2.11) is a sufficient condition for the entry equilibrium to exist. But if the candidate offer is not incentive-compatible, then the entrant will have to make an alternative offer, which will generate less than the maximal total surplus $CS_E^l(c_E) + mCS_E^s(c_E)$, which means that the feasibility condition tightens.

Proposition 11 (*entry under rebates*)

(i) *If E's joint net surplus is not sufficient to match both I's best offer to the large buyer and I's best offer to the small buyers simultaneously, then no "entry equilibrium" exists, even if buyers buy from E whenever E's offers to each of them are at least as good as I's offers.*

(ii) *If*

- E 's joint net surplus is sufficient to match both I 's best offer to the large buyer and I 's best offer to the small buyers simultaneously, and

- E 's offer satisfies the large buyer's "self-selection condition" of equation (2.10),

then our game has a pure-strategy equilibrium where all buyers buy from E after observing that E 's offers to each of them are at least as good as I 's offers. ("entry equilibrium"). Such an equilibrium can arise even when $c_I < 1 - \sqrt{1/2}$, i.e. when there would not be an entry equilibrium under third-degree price discrimination.

Proof: see Appendix A

We see that even uniform rebate schemes will allow the incumbent to strategically redistribute rents between different types of buyers so as to prevent the more efficient entrant from serving the buyers. Thus, if parameters are such that no entry equilibrium exists even under the continuation equilibria specified in Proposition 11 (which are in favor of the entrant), then the miscoordination equilibrium characterized in Lemma 10 is the only pure-strategy equilibrium of our game. Whenever the miscoordination equilibrium is unique, we refer to it as "exclusionary equilibrium".

If instead parameters are such that the entry equilibrium exists, then our game has two pure-strategy equilibria: one where all buyers buy from the entrant, and one where they all buy from the incumbent (miscoordination equilibrium). Then, it depends on the continuation equilibria which of these two types of equilibria will be played.

There is an interesting general pattern that emerges from both Lemma 10 and Proposition 11: If the self-selection constraint is not binding for either type of buyer, both large

and small buyers are charged the same unit price. If instead the large buyer's self-selection constraint is binding, the firm will raise the price of the small buyers *above* c_i (while $p_i^l = c_i$); if it is the small buyers' self-selection constraint that is binding, the firm will lower the large buyer's price *below* c_i , while $p_i^s = c_i$ (unless of course $c_i = 0$, as is the case for the entrant, who cannot charge a price below marginal cost because we ruled out negative prices).

In other words, whenever p_i^s differs from p_i^l , we have that $p_i^s > p_i^l$. But that seems like a fairly typical feature of real-life price schedules. What is interesting is that this price pattern has nothing to do with decreasing marginal cost or the like, but it arises out of the self-selection constraint implied by second-degree price discrimination.

Welfare Analysis

We have shown that under uniform linear prices, the incumbent cannot prevent entry when buyers coordinate on the entrant (no matter how small the efficiency gap between I and E), while rebate schemes can indeed be designed so as to break entry even if the entrant is significantly more efficient than the incumbent. The exclusionary potential of two-part tariffs is even greater if these tariffs are allowed to depend on the type of buyer (not just the quantity they buy).

But the price regime does not only affect the likelihood of entry, it also determines the distribution of rents among firms and buyers and the size of possible efficiency losses under those equilibria that exist for a particular type of price regime.

Table 1 shows the total surplus of large and small buyers, the profits of incumbent and

entrant, as well as the allocative and productive efficiency loss under each price regime, depending on whether the equilibrium has the incumbent or the entrant serve the buyers. Note that the first best allocation has E serve all buyers at marginal cost $c_E = 0$, which generates total surplus $1/2$, so that the welfare loss under any alternative allocation is the total surplus generated by the allocation under consideration minus $1/2$.

Given that the first-best outcome has the entrant serve the buyers, it is obvious that whenever the incumbent serves, the equilibrium will be inefficient, because the incumbent produces at a higher marginal cost than the entrant. The resulting efficiency losses will be smallest under third-degree price discrimination, and largest under uniform linear prices, while they are in-between under uniform rebates.⁸

Full efficiency can only arise if the entrant serves, and if the entrant can use two-part tariffs⁹, where it is irrelevant if these tariffs are discriminatory or uniform. The smallest possible welfare loss when I serves (namely $-\frac{1}{2} + \frac{1}{2}(1 - c_I)^2$) is exactly equal to the largest possible welfare loss when E serves. Of course, when E serves, there is no productive inefficiency, so all remaining welfare losses must be allocative.¹⁰

⁸Note that the efficiency losses under third-degree price discrimination also represent the lower bound on efficiency losses under any alternative equilibrium (no matter which price regime) where the incumbent cannot fully exploit the buyers because the continuation equilibria are less favorable to the incumbent (e.g. buyers will only miscoordinate on the incumbent if the latter charges at most $p_I = c_I$ to all buyers).

⁹unless $\bar{s} = 1$, in which case even linear prices would yield full efficiency (because the entrant would have to set price equal to marginal cost: $p_E = 1 - \bar{s} = 0$)

¹⁰Note that the efficiency losses under uniform linear prices also represent the upper bound on efficiency losses under any alternative equilibrium (no matter which price regime) where the continuation equilibria are less favorable to the entrant (e.g. buyers will only coordinate on the entrant if p_E is strictly less than $\min\{c_I, 1 - \bar{s}\}$).

Table 1: Buyers' Surplus, Profits, and Welfare Loss

(a) Buyers' Net Total Surplus			
Price Regime	Small Buyers	Large Buyer	Total
Linear Prices			
<i>I</i> serves	$\frac{1}{8} \frac{K}{m} (1 - c_I)^2$	$\frac{1}{8} (1 - K) (1 - c_I)^2$	$\frac{1}{8} (1 - c_I)^2$
<i>E</i> serves	$\frac{K}{2m} \max \{ (1 - c_I)^2, \bar{s}^2 \}$	$\frac{1-K}{2} \max \{ (1 - c_I)^2, \bar{s}^2 \}$	$\frac{1}{2} \max \{ (1 - c_I)^2, \bar{s}^2 \}$
3 rd -degree PD			
<i>I</i> serves	0	0	0
<i>E</i> serves	$\frac{1}{m} \frac{1}{2} (1 - c_I)^2$	$\frac{1}{2} (1 - c_I)^2$	$(1 - c_I)^2$
2 nd -degree PD			
<i>I</i> serves	0	$\frac{K(1-p_I^*(c_I))^2}{2m} \left(1 - \frac{K}{m} \frac{1}{1-K}\right)$	$\frac{K(1-p_I^*(c_I))^2}{2m} \left(1 - \frac{K}{m} \frac{1}{1-K}\right)$
<i>E</i> serves	$< \frac{1}{m} \frac{1}{2} (1 - c_I)^2$	$\leq \frac{1}{2} (1 - c_I)^2$	$< (1 - c_I)^2$
(b) Firms' Profits			
Price Regime	Incumbent	Entrant	Total
Linear Prices			
<i>I</i> serves	$\frac{1}{4} (1 - c_I)^2$	0	$\frac{1}{4} (1 - c_I)^2$
<i>E</i> serves	0	$c_I (1 - c_I)$ or $(1 - \bar{s}) \bar{s}$	$c_I (1 - c_I)$ or $(1 - \bar{s}) \bar{s}$
3 rd -degree PD			
<i>I</i> serves	$\frac{1}{2} (1 - c_I)^2$	0	$\frac{1}{2} (1 - c_I)^2$
<i>E</i> serves	0	$\frac{1}{2} - (1 - c_I)^2$	$\frac{1}{2} - (1 - c_I)^2$
2 nd -degree PD			
<i>I</i> serves	$\in \left(\frac{(1-c_I)^2}{4}, \frac{(1-c_I)^2}{2} \right)$	0	$\in \left(\frac{(1-c_I)^2}{4}, \frac{(1-c_I)^2}{2} \right)$
<i>E</i> serves	0	$> \frac{1}{2} - (1 - c_I)^2$	$> \frac{1}{2} - (1 - c_I)^2$

(c) Welfare Loss			
Price Regime	Allocative	Productive	Total
Linear Prices			
<i>I</i> serves	$-\frac{1}{2} \left(\frac{1+c_I}{2} \right)^2$	$-c_I \frac{1}{2} (1 - c_I)$	$-\frac{1}{2} + \frac{1}{8} (1 - c_I)^2$
<i>E</i> serves	$-\frac{1}{2} + \max \left\{ \frac{(1-c_I)^2}{2}, \frac{s^2}{2} \right\}$	0	$-\frac{1}{2} + \max \left\{ \frac{(1-c_I)^2}{2}, \frac{s^2}{2} \right\}$
3 rd -degree PD			
<i>I</i> serves	$-\frac{c_I^2}{2}$	$-c_I (1 - c_I)$	$-\frac{1}{2} + \frac{1}{2} (1 - c_I)^2$
<i>E</i> serves	0	0	0
2 nd -degree PD			
<i>I</i> serves	$-\frac{(1-K)c_I^2}{2} - \frac{K(p_I^s(c_I))^2}{2}$	$-c_I (1 - K) (1 - c_I) -$ $-c_I K (1 - p_I^s(c_I))$	$\frac{(1-c_I)^2}{2} - \frac{1}{2} -$ $-K (p_I^s(c_I) - c_I)^2$
<i>E</i> serves	0	0	0

We can conclude that there is a trade-off between maximizing *E*'s chances to enter, and minimizing welfare losses (no matter which firm eventually serves the buyers). Discriminatory two-part tariffs raise the highest barriers to *E*'s entry, but yield the most efficient outcome (full efficiency if *E* serves, lowest-possible inefficiency if *I* serves). The opposite is true for linear tariffs (lowest entry barriers, but least efficient outcomes).

Uniform rebates are somewhere between these two extremes. Note, however, that they are sufficient to achieve full efficiency. In other words, if *E* is sufficiently efficient to enter under uniform rebates, then allowing discriminatory two-part tariffs will not yield any

efficiency gains, but may jeopardize E 's entry.¹¹

On the other hand, banning two-part tariffs altogether will only make sense if the efficiency gap between I and E is small, so that even uniform rebates represent a serious barrier to E 's entry. But in this case, the welfare gains that can be expected from E 's entry are low anyhow. Moreover, having E serve the buyers under linear prices will yield (almost or exactly) the same surplus as having I serve the buyers under discriminatory two-part tariffs.

The conclusions are somewhat different if the welfare criterion is buyers' surplus, not social efficiency. We see that if the incumbent serves, both large and small buyers will prefer linear prices over any type of two-part tariffs. If instead the entrant serves, the small buyers will always prefer discriminatory two-part tariffs over uniform rebates, while the large buyer is indifferent between the two.

If $1 - c_I > \bar{s}$, then both types of buyers strictly prefer two-part tariffs over linear prices. If instead $1 - c_I < \bar{s}$, then the ranking is ambiguous. The small buyers will prefer discriminatory two-part tariffs over linear prices iff $K < (1 - c_I/\bar{s})^2$, and the same holds for the large buyer iff $1 - K < (1 - c_I/\bar{s})^2$.

Concluding remarks

The purpose of this exercise was to demonstrate the exclusionary potential of rebate arrangements in the presence of network externalities. We have shown that two-part tariffs, even if required to be non-discriminatory, may allow the incumbent to prevent entry of a

¹¹Discriminatory two-part tariffs only make sense if one can take for granted that buyers will miscoordinate on the incumbent.

more efficient firm in cases where that would not have been possible under uniform flat prices. The finding is particularly interesting insofar as, in our model, the entrant is in a fairly good initial position compared to other papers on exclusionary practices: it does not have to pay any fixed cost to start operating in the industry, entrant and incumbent can approach buyers simultaneously (i.e. the incumbent has no first-mover advantage in offering contracts to the buyers before the entrant can do so), and the entrant has the same pricing instruments at its disposal.

Our analysis is very preliminary and should be further developed in the following directions:

- What if buyers can coordinate their actions among each other?
- What if buyers are allowed to patronize more than one firm?
- Is our model robust to the large buyer being sufficient for the entrant to reach the minimum size?

Interesting extensions of our model could be to allow for buyers to compete against each other downstream, to see whether the same kind of results as in Fumagalli and Motta (2005) would arise. Another issue of interest could be to allow for partial (or even full) compatibility between I 's and E 's network, and to introduce compatibility as a strategic choice variable.

Finally, note that we should expect very similar results in a model where there are no network externalities, but instead buyers have switching costs, and the entrant faces some fixed costs of entry. Farrell and Klemperer (forthcoming) have pointed out the analogies

between network externalities and switching costs:

"Both switching costs and proprietary network effects arise when consumers value forms of *compatibility* that require otherwise separate purchases to be made from the same firm. Switching costs arise if a consumer wants a group, or especially a series, of his own purchases to be compatible with one another: this creates economies of scope among his purchases from a single firm. Network effects arise when a user wants compatibility with *other* users (or complementors), so that he can interact or trade with them, or use the same complements; this creates economies of scope between *different* users' purchases.

These economies of scope make it unhelpful to isolate a transaction: a buyer's best action depends on other, complementary transactions. When those transactions are in the future, or made simultaneously by others, his *expectations* about them are crucial. When they are in the past, they are *history* that matters to him. History also matters to a firm because established *market share* is a valuable asset: in the case of switching costs, it represents a stock of individually locked-in buyers, while in the case of network effects an installed base directly lets the firm offer more network benefits and may also boost expectations about future sales." (Farrell and Klemperer (2004): page 1 - Introduction)

Bibliography

- [1] AGHION, PHILIPPE/BOLTON, PATRICK (1987): "Contracts as a Barrier to Entry", *American Economic Review* 77 (3), 388 - 401.
- [2] BERNHEIM, B. DOUGLAS/WHINSTON, MICHAEL D. (1998): "Exclusive Dealing", *Journal of Political Economy* 106 (1), 64 - 103.
- [3] FARRELL, JOE/KLEMPERER, PAUL (2004): "Coordination and Lock-In: Competition with Switching Costs and Network Effects", preliminary draft chapter for: *Handbook of Industrial Organization*, Vol. 3, Joe Farrell and Paul Klemperer (forthcoming), http://www.nuff.ox.ac.uk/users/klemperer/SectionI_17Dec.pdf.
- [4] FUMAGALLI, CHIARA/MOTTA, MASSIMO (2005): "Exclusive Dealing and Entry When Buyers Compete", *American Economic Review*, forthcoming.
- [5] GYSELEN, LUC (2003): "Rebates: Competition on the Merits or Exclusionary Practice?", in Ehlermann, Claus-Dieter / Atanasiu, Isabela (eds): *European Competition Law Annual 2003: What Is an Abuse of a Dominant Position?*, Oxford/Portland Oregon, Hart Publishing.
- [6] INNES, ROBERT/SEXTON, RICHARD J. (1993): "Customer coalitions, monopoly price discrimination and generic entry deterrence", *European Economic Review* 37 (1993), 1569 - 1597.
- [7] INNES, ROBERT/SEXTON, RICHARD J. (1994): "Strategic Buyers and Exclusionary Contracts", *American Economic Review* 84 (3), 566 - 584.

- [8] RASMUSEN, ERIC B./RAMSEYER, J. MARK/WILEY, JOHN S. (1991): "Naked Exclusion", *American Economic Review* 81 (5), 1137 - 45.
- [9] SEGAL, ILYA R./WHINSTON, MICHAEL D. (2000): "Naked exclusion: Comment", *American Economic Review* 90 (1), 296 - 309.

Appendix A: Proofs

Proof of Proposition 7:

(i) Let $\bar{s} \leq 1 - c_I$. Then, with all buyers buying from E at $p_E = c_I$, total demand is $mq_E^s(p_E) + q_E^l(p_E) = 1 - c_I \geq \bar{s}$, and so E will reach the minimum size. Thus, E 's product has the exact same value to the buyers as I 's, and it sells at the same price, so that buyers are indifferent between I 's and E 's offer. I will not want to deviate either: To attract the buyers, I would have to set a price $p_I < c_I$, i.e. sell at a loss; and increasing p_I above c_I will not attract any buyers. E has no incentive to change anything about its price either: increasing p_E would imply losing the buyers to I , and decreasing p_E will just reduce profits (recall that E is not radically more efficient than I , i.e. $p_E = c_I < p_E^*$, so that E cannot gain from reducing its price below c_I).

Let $\bar{s} > 1 - c_I$. Then, total demand at $p_E = 1 - \bar{s}$ is $mq_E^s(p_E) + q_E^l(p_E) = 1 - p_E = \bar{s}$, and so E will just reach the minimum size, while still breaking even ($\bar{s} \leq 1$ and $c_E = 0$ imply $p_E - c_E \geq 0$). E has no incentive to increase its price, as that would imply falling short of the minimum size (followed by a break-down of coordination, i.e. all buyers would buy from I), while charging a price below $1 - \bar{s}$ would only reduce profits. The buyers have no incentive to deviate and buy from I , because $p_I > p_E$. If I decreases its price to a value

$p_I \in (1 - \bar{s}, p_I^*)$ or increases it to some $p_I > p_I^*$, I will not be able to attract any buyers, and selling at a price at or below $1 - \bar{s}$ (which is strictly smaller than c_I if $\bar{s} > 1 - c_I$) would imply losses.

Note that we eliminate all equilibria in weakly dominated strategies, where

- if $\bar{s} \leq 1 - c_I$, I sets $p_I \in [0, c_I]$ instead of $p_I = c_I$, and E sets $p_E = p_I$, and
- if $\bar{s} > 1 - c_I$, I sets $p_I \in [1 - \bar{s}, p_I^*)$ or $p_I > p_I^*$, and E sets $p_E = 1 - \bar{s}$.

(ii) Suppose that all buyers buy from I . Then, recall that $\bar{s} > \max \{1 - K, K\}$, implying that none of the individual buyers alone is sufficient for E to reach the minimum size. Thus, E 's product has zero value for any single buyer, and so no buyer will want to deviate and buy from E , even though $p_I > p_E$. I sets $p_I = p_I^*$, which is the most profitable among all prices $p_I \leq p_I^* + p_E^*$ under which buyers will miscoordinate on the incumbent. Thus, I has no incentive to increase or decrease its price. Since buyers will not switch to E even if the price difference between the two firms is maximal, i.e. even if E charges $p_E = 0$ (so that $p_I = p_I^* + p_E$), E has no incentive to decrease its price.

We eliminate all equilibria in weakly dominated strategies, where I sets $p_I = p_I^*$, and E sets $p_E \neq p_E^*$. \square

Proof of Lemma 10:

We will first show that I 's equilibrium offer coincides with the solution to the following profit maximization problem:

$$\max_{\{p_I^s, p_I^l, R_I^s, R_I^l\}} m(p_I^s - c_I) q_I^s(p_I^s) + m(-R_I^s) + (p_I^l - c_I) q_I^l(p_I^l) + (-R_I^l)$$

subject to the large and small buyers' self-selection constraints, the large and small buyers' participation constraints, and the break-even constraint:

- (i) $CS_I^l(p_I^l) + R_I^l \geq CS^{l,net}(p_I^s, q_I^s(p_I^s)) + R_I^s$
- (ii) $CS_I^s(p_I^s) + R_I^s \geq CS^{s,net}(p_I^l, q_I^l(p_I^l)) + R_I^l$
- (iii) $CS_I^l(p_I^l) + R_I^l \geq 0$
- (iv) $CS_I^s(p_I^s) + R_I^s \geq 0$
- (v) $m(p_I^s - c_I)q_I^s(p_I^s) + (p_I^l - c_I)q_I^l(p_I^l) \geq mR_I^s + R_I^l$

Now, at the optimum, constraints (i) and (iv) will be binding, thus determining R_I^l and R_I^s . This, in turn, implies that $p_I^l = c_I$ (which, given p_I^s , maximizes the rent that I can extract from the large buyer).

Therefore, the incumbent's problem reduces to choosing the right p_I^s . The resulting maximization problem is convex in p_I^s , and solves for

$$p_I^{s,opt} = \frac{c_I + \frac{1}{m} \left(1 - \frac{1}{1-K} \frac{K}{m}\right)}{1 + \frac{1}{m} \left(1 - \frac{1}{1-K} \frac{K}{m}\right)}$$

Our assumption that $c_I < 1$ implies $p_I^{s,opt} \in (c_I, 1)$.

The large buyer's participation constraint (iii) will be oversatisfied under this solution:

Given that constraints (i) and (iv) hold with equality, we have

$$CS_I^l(c_I) + R_I^l = CS^{l,net}(p_I^{s,opt}, q_I^s(p_I^{s,opt})) - CS_I^s(p_I^{s,opt})$$

But the right-hand side of this equality is strictly positive because

$$CS^{l,net}(p_I^s, q_I^s(p_I^s)) > CS_I^s(p_I^s) \text{ for all } p_I^s$$

which implies that $CS_I^l(c_I) + R_I^{l,opt} > 0$.

The small buyers' self-selection condition (ii) holds as well: Suppose a small buyer considers buying the large buyer's sales target, $\bar{q}_I^l = q_I^l(c_I)$, at price c_I . Then, this buyer would enjoy net consumer surplus $CS_I^s(c_I)$ from consuming the first $q_I^s(c_I)$ units, and negative surplus on all the remaining units $\bar{q}_I^l - q_I^s(c_I)$ (which have to be bought at price c_I , but have a value less than c_I to the small buyer). Thus, the small buyer's net consumer surplus from consuming $\bar{q}_I^l = q_I^l(c_I)$ will be strictly lower than $CS_I^s(c_I)$. However, the small buyer will have to pay $R_I^{l,opt}$ (instead of $R_I^s = -CS_I^s(p_I^{s,opt})$), where $R_I^{l,opt} < -CS_I^s(c_I)$. Thus, a small buyer buying the large buyer's threshold would end up with a strictly negative net total surplus, and so he will prefer to buy the small buyers' threshold $\bar{q}_I^s = q_I^s(p_I^{s,opt})$.

Finally, I 's break-even constraint (v) reduces to

$$m(p_I^{s,opt} - c_I)q_I^s(p_I^{s,opt}) + 0 \geq mR_I^s + R_I^l$$

which holds because both large and small buyers will pay strictly positive fees to the incumbent ($R_I^s < 0$ and $R_I^l < 0$).

Hence, the incumbent's equilibrium offer

$$\begin{aligned} p_I^s &= p_I^{s,opt} > c_I, p_I^l = c_I, \bar{q}_I^s = q_I^s(p_I^{s,opt}), \bar{q}_I^l = q_I^l(c_I) \\ R_I^s &= -CS_I^s(p_I^{s,opt}), R_I^l = CS_I^{l,net}(p_I^{s,opt}, q_I^s(p_I^{s,opt})) + R_I^s - CS_I^l(c_I) \end{aligned}$$

is the most profitable among all feasible offers, and so I will not have any incentive to deviate.

Given that buyers miscoordinate on the incumbent no matter which offer the entrant makes, the entrant is indifferent between all feasible offers (the reasoning is analogous to the Proof of Lemma 8). Eliminating all equilibria where E plays weakly dominated strategies,

E will just solve the analogous optimization problem analyzed above, which yields

$$\begin{aligned} p_E^{s,opt} &= \frac{\frac{1}{m} \left(1 - \frac{1}{1-K} \frac{K}{m}\right)}{1 + \frac{1}{m} \left(1 - \frac{1}{1-K} \frac{K}{m}\right)} < p_I^{s,opt} \\ p_E^s &= p_E^{s,opt} > 0, p_E^l = c_E = 0 < p_I^l, \bar{q}_{E,s} = q_E^s(p_E^{s,opt}), \bar{q}_{E,l} = q_E^l(c_E) \\ R_E^s &= -CS_E^s(p_E^{s,opt}), R_E^l = CS_E^{l,net}(p_E^{s,opt}, q_E^s(p_E^{s,opt})) + R_E^s - CS_E^l(c_E) \end{aligned}$$

But given that all buyers buy from I , no individual buyer will want to deviate and buy from E , and so all buyers will end up buying from I . \square

Proof of Proposition 11:

(i) We argued in Lemma 10 that miscoordination equilibria exist for all parameter values under the appropriate continuation equilibria. We will now show that entry equilibria will only exist for certain parameter values, even under those continuation equilibria that are "favorable" to the entrant.

Analogously to Proposition 9, a necessary condition for existence of an entry equilibrium is that I can neither match E 's offer to the large buyer nor E 's offer to the small buyers.

The best offer that I can make to the small buyers solves

$$\max_{\{p_I^s, p_I^l, R_I^s, R_I^l\}} CS_I^s(p_I^s) + R_I^s$$

subject to the large and small buyers' self-selection constraints, the large and small buyers' participation constraints, and the break-even constraint:

- (i) $CS_I^l(p_I^l) + R_I^l \geq CS^{l,net}(p_I^s, q_I^s(p_I^s)) + R_I^s$
- (ii) $CS_I^s(p_I^s) + R_I^s \geq CS^{s,net}(p_I^l, q_I^l(p_I^l)) + R_I^l$
- (iii) $CS_I^l(p_I^l) + R_I^l \geq 0$
- (iv) $CS_I^s(p_I^s) + R_I^s \geq 0$
- (v) $m(p_I^s - c_I)q_I^s(p_I^s) + (p_I^l - c_I)q_I^l(p_I^l) \geq mR_I^s + R_I^l$

Now, at the optimum, constraints (i) and (v) will be binding, thus determining R_I^l and R_I^s , and that $p_I^l = c_I$ (which, given p_I^s , maximizes the rent that I can extract from the large buyer).

Thus, I 's reduced problem is convex in p_I^s , and solves for

$$p_I^{s,best} = \frac{\frac{1}{m+1} \left(1 + c_I + \frac{2}{m} \left(1 - \frac{1}{2} \frac{1-K}{1-K} \frac{K}{m} \right) \right) - \frac{1}{m}}{\frac{1}{m+1} \left(2 + \frac{2}{m} \left(1 - \frac{1}{2} \frac{1-K}{1-K} \frac{K}{m} \right) \right) - \frac{1}{m}}$$

where $p_I^{s,best} \in (c_I, 1)$. The small buyers' self-selection condition (ii) reduces to

$$CS_I^l(c_I) - CS^{l,net}(p_I^{s,best}, q_I^s(p_I^{s,best})) \geq CS^{s,net}(c_I, q_I^l(c_I)) - CS_I^s(p_I^{s,best})$$

which holds for all $p_I^s \geq c_I$. The participation constraints (iii) and (iv) are over-satisfied at the solution:

$$CS_I^s(p_I^{s,best}) + R_I^{s,best} > 0 \text{ and } CS_I^l(c_I) + R_I^{l,best} > 0.$$

Note that this solution will necessarily yield lower total surplus for the small buyers than the corresponding offer under third-degree price discrimination, where each buyer receives $CS_I^s(c_I) + CS_I^l(c_I)/m$.

The incumbent's best offer to the large buyer solves the analogous program:

$$\max_{\{p_I^s, p_I^l, R_I^s, R_I^l\}} CS_I^l(p_I^l) + R_I^l$$

subject to the large and small buyers' self-selection constraints, the large and small buyers' participation constraints, and the break-even constraint:

- (i) $CS_I^l(p_I^l) + R_I^l \geq CS^{l,net}(p_I^s, q_I^s(p_I^s)) + R_I^s$
- (ii) $CS_I^s(p_I^s) + R_I^s \geq CS^{s,net}(p_I^l, q_I^l(p_I^l)) + R_I^l$
- (iii) $CS_I^l(p_I^l) + R_I^l \geq 0$
- (iv) $CS_I^s(p_I^s) + R_I^s \geq 0$
- (v) $m(p_I^s - c_I)q_I^s(p_I^s) + (p_I^l - c_I)q_I^l(p_I^l) \geq mR_I^s + R_I^l$

This problem has four different solutions:

Case 1: $0 < K \leq \frac{c_I(1-c_I)}{\frac{1}{2}\frac{1}{m} + c_I(1-c_I) + \frac{1}{2}(1-c_I)^2} = K^*(c_I, m)$

Only constraints (iv) and (v) are binding, all other constraints are oversatisfied. I can extract the full net consumer surplus from the small buyers and transfer it to the large buyer without violating the small buyers' self-selection constraint (ii):

$$\begin{aligned} p_I^{s,best} &= p_I^{l,best} = c_I, \bar{q}_I = q_I^l(c_I) \\ R_I^{s,best} &= -CS_I^s(c_I), R_I^{l,best} = mCS_I^s(c_I) \end{aligned}$$

Case 2: $K^*(c_I, m) < K \leq \frac{c_I(1-\frac{1}{m+1}c_I)}{\frac{1}{2}\frac{1}{m} + c_I(1-\frac{1}{m+1}c_I) + \frac{1}{2}(1-c_I)^2} = K^{**}(c_I, m)$

Constraints (ii), (iv), and (v) are binding, while all other constraints are over-satisfied.

The solution reads

$$\begin{aligned} p_I^{s,best} &= c_I, p_I^{l,best} = 1 - \frac{1}{2} \frac{1}{c_I} \frac{K}{m} \frac{1}{1-K} (1 + m(1-c_I)^2) \\ R_I^{s,best} &= -CS_I^s(c_I), R_I^{l,best} = -\frac{1}{2} \frac{K}{m} + p_I^{l,best} (1 - p_I^{l,best}) (1-K) \end{aligned}$$

where $K > K^*(c_I, m)$ implies that $p_I^{l,best} < c_I$, and $K \leq K^{**}(c_I, m)$ implies $p_I^{l,best} > 0$.

Case 3: $K^{**}(c_I, m) < K \leq \frac{m+1-c_I}{m+2+\frac{1}{m}-c_I} = K^{***}(c_I, m)$

Constraints (ii) and (v) are binding, while all other constraints are over-satisfied. The solution reads

$$\begin{aligned} p_I^{s,best} &= c_I, p_I^{l,best} = \frac{1}{m+1} c_I \\ R_I^{s,best} &= \frac{1}{m+1} \left(\frac{1}{2} \frac{K}{m} - \frac{1}{2} \frac{K}{m} (1-c_I)^2 - c_I (1-K) \left(1 - \frac{1}{m+1} c_I \right) \right) \\ R_I^{l,best} &= -\frac{1}{m+1} \frac{K}{2} c_I (2-c_I) \end{aligned}$$

Case 4: $K^{***}(c_I, m) < K < \frac{m}{m+1}$

This case is analogous to Case 3, i.e. constraints (ii) and (v) are binding, but now the small buyers are fairly large $q_I^s(p_I^{l,best}) \leq \frac{K}{m}$, and so the solution reads:

$$\begin{aligned} p_I^{s,best} &= c_I, p_I^{l,best} = \frac{\frac{m}{K}(1-K) - 1 - \frac{c_I}{m}}{\frac{m}{K}(1-K) - 1 - \frac{1}{m}} \\ R_I^{s,best} &= \frac{1}{m+1} \left(\frac{(1-c_I)^2 (1-K) (-\frac{1}{m})}{\left(\frac{m}{K}(1-K) - 1 - \frac{1}{m} \right)^2} \left(\frac{m}{K}(1-K) - 1 - \frac{1}{m} + \frac{1}{2} \frac{1-K}{K} \right) - \frac{1}{2} \frac{K}{m} (1-c_I)^2 \right) \\ R_I^{l,best} &= \frac{1}{2} \frac{1}{m+1} (1-c_I)^2 \left[K - \frac{(1-K)^2}{K \left(\frac{m}{K}(1-K) - 1 - \frac{1}{m} \right)^2} \right] \end{aligned}$$

This problem is not convex in p_I^l ; the second-order condition requires $\frac{m}{K}(1-K) - 1 - \frac{1}{m} < 0$; however, $K > K^{***}(c_I, m)$ implies that the second-order condition is satisfied and that $p_I^{l,best} > 0$.

(There is another solution where $p_I^{s,best} = c_I$, $p_I^{l,best} = 0$, $R_I^{l,best} = -\frac{1}{2} \frac{K}{m}$, $R_I^{s,best} = -CS_I^s(c_I)$. This offer is feasible iff $K \geq c_I / (\frac{1}{2} \frac{1}{m} + c_I + \frac{1}{2} (1-c_I)^2)$. But this solution is just a local maximum; it yields less total surplus for the large buyer than the solutions of case 3 and 4, which are the alternative solutions for the relevant range of K).

Note that in cases 2, 3, and 4, we have $p_I^{l,best} \in (0, c_I)$, i.e. I sells to the large buyer

below marginal cost (thus incurring losses), and the large buyer's total net surplus will be lower than under the "first best" offer, where it is $CS_I^l(c_I) + mCS_I^s(c_I)$.

Now, if the total net surplus that E can generate, $CS_E^l(c_E) + mCS_E^s(c_E)$, is smaller than the sum of I 's best offer to the large buyer and I 's best offer to the small buyers, then I can always match either E 's offer to the large buyer, or E 's offer to the small buyers, implying that there is no entry equilibrium in pure strategies.

The corresponding "minimum efficiency" condition reads:

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq CS_I^l(p_I^{l,best}) + R_I^{l,best} + m \left(CS_I^s(p_I^{s,best}) + R_I^{s,best} \right)$$

which completes the proof.

(ii) We argued in Lemma 10 that our game always has an equilibrium where all buyers buy from I under the appropriate coordination equilibria.

For an entry equilibrium to exist, E 's offer must satisfy the following conditions:

- it matches I 's best offer to both the large and the small buyers;
- it satisfies the large and small buyers' self-selection condition;
- E reaches the minimum size; and
- E breaks even.

More formally, E has to solve the following program:

$$\max_{\{p_E^s, p_E^l, R_E^s, R_E^l\}} m(p_E^s - c_E) q_E^s(p_E^s) + (p_E^l - c_E) q_E^l(p_E^l) - mR_E^s - R_E^l$$

subject to:

- (i) $CS_E^l(p_E^l) + R_E^l \geq CS_I^l(p_I^{l,best}) + R_I^{l,best}$
- (ii) $CS_E^s(p_E^s) + R_E^s \geq CS_I^s(p_I^{s,best}) + R_I^{s,best}$
- (iii) $CS_E^l(p_E^l) + R_E^l \geq CS^{l,net}(p_E^s, q_E^s(p_E^s)) + R_E^s$
- (iv) $CS_E^s(p_E^s) + R_E^s \geq CS^{s,net}(p_E^l, q_E^l(p_E^l)) + R_E^l$
- (v) $q_E^l(p_E^l) + mq_E^s(p_E^s) \geq \bar{s}$
- (vi) $m(p_E^s - c_E)q_E^s(p_E^s) + (p_E^l - c_E)q_E^l(p_E^l) \geq mR_E^s + R_E^l$

Case 1: Let the total surplus generated by E be larger than the sum of I 's best offer to the large and small buyers:

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq CS_I^l(p_I^{l,best}) + R_I^{l,best} + m(CS_I^s(p_I^{s,best}) + R_I^{s,best})$$

and let the large buyer's self-selection condition be satisfied under the "first-best" solution (defined below):

$$CS_I^l(p_I^{l,best}) + R_I^{l,best} - (CS_I^s(p_I^{s,best}) + R_I^{s,best}) \geq CS^{l,net}(c_E, q_E^s(c_E)) - CS_E^s(c_E)$$

Then, the solution to E 's problem is ("first-best" solution):

$$\begin{aligned} p_E^{s*} &= p_E^{l*} = c_E = 0 \\ R_E^{l*} &= CS_I^l(p_I^{l,best}) + R_I^{l,best} - CS_E^l(c_E) \\ R_E^{s*} &= CS_I^s(p_I^{s,best}) + R_I^{s,best} - CS_E^s(c_E) \end{aligned}$$

Under this solution, constraints (i) and (ii) are satisfied (they hold with equality) by construction of R_E^{l*} and R_E^{s*} . Constraint (iii) holds by assumption. Constraint (v) holds by condition (2.4):

$$mq_E^s(c_E) + q_E^l(c_E) = 1 \geq \bar{s}$$

Constraint (vi) reduces to

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq CS_I^l(p_I^{l,best}) + R_I^{l,best} + m \left(CS_I^s(p_I^{s,best}) + R_I^{s,best} \right)$$

which is satisfied by assumption.

We will now argue that if constraints (i) and (ii) are satisfied with strict equality and $p_E^l = c_E = 0$, then constraint (iv) is satisfied as well, i.e. the small buyers' self-selection constraint is redundant: First of all,

$$CS_E^s(c_E) + R_E^{s*} = CS_I^s(p_I^{s,best}) + R_I^{s,best}$$

by strict equality of (ii), and

$$CS_I^s(p_I^{s,best}) + R_I^{s,best} \geq CS_I^s(c_I) + R_I^{s,best}$$

by optimality of $p_I^{s,best} \in (c_I, 1)$ among all other feasible pairs (p_I^s, R_I^s) , and in particular the pair $(c_I, R_I^{s,best})$, where

$$R_I^{s,best} = \max \left\{ -CS_I^s(c_I), CS^{s,net} \left(p_I^{l,best}, q_I^l(p_I^{l,best}) \right) + R_I^{l,best} - CS_I^s(c_I) \right\}$$

is the fixed payment that the small buyers have to make to I under I 's best offer to the large buyer. Then, by definition of $R_I^{s,best}$, we have

$$CS_I^s(c_I) + R_I^{s,best} \geq CS^{s,net} \left(p_I^{l,best}, q_I^l(p_I^{l,best}) \right) + R_I^{l,best}$$

i.e. I 's best offer to the large buyer satisfies the small buyers' self-selection condition. Finally, we need to show that

$$CS^{s,net} \left(p_I^{l,best}, q_I^l(p_I^{l,best}) \right) + R_I^{l,best} \geq CS^{s,net} (c_E, q_E^l(c_E)) + R_E^{l*}$$

Since constraint (i) holds with equality, we have that $R_I^{l,best} - R_E^{l*} = CS_E^l(c_E) - CS_I^l(p_I^{l,best})$, so that the above inequality can be rearranged to read

$$CS^{s,net}(p_I^{l,best}, q_I^l(p_I^{l,best})) - CS^{s,net}(c_E, q_E^l(c_E)) + (CS_E^l(c_E) - CS_I^l(p_I^{l,best})) \geq 0$$

If $q_I^l(p_I^{l,best}) \geq \frac{K}{m}$, we can insert for the consumer surplus terms from equations (2.6) and (2.9) and reduce the inequality to $p_I^{l,best} \geq c_E = 0$, which is always true. If instead $q_I^l(p_I^{l,best}) < \frac{K}{m}$, inserting the appropriate terms yields

$$1 - K \geq \frac{K}{m} + (1 - K) \left(1 - p_I^{l,best}\right)^2 \left((1 - K) \frac{K}{m} - 1\right)$$

which holds as well, because the LHS is strictly larger than $\frac{K}{m}$, while the RHS is strictly smaller than $\frac{K}{m}$. Thus, we can conclude that the small buyers' self-selection condition is implied to hold under our candidate solution:

$$CS_E^s(c_E) + R_E^{s*} \geq CS^{s,net}(c_E, q_E^l(c_E)) + R_E^{l*}$$

The candidate solution satisfies all constraints (i) to (vi), and $p_E^s = p_E^l = c_E = 0$ maximizes consumer surplus, and hence the total surplus that E can appropriate.

Finally, recall that we argued above that $CS_I^l(p_I^{l,best}) + R_I^{l,best} \leq CS_I^l(c_I) + mCS_I^s(c_I)$, and that $m(CS_I^s(p_I^{s,best}) + R_I^{s,best}) < CS_I^l(c_I) + mCS_I^s(c_I)$. Thus, the candidate entry equilibrium will arise whenever

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq CS_I^l(p_I^{l,best}) + R_I^{l,best} + m(CS_I^s(p_I^{s,best}) + R_I^{s,best})$$

Comparing this inequality to the corresponding condition under third degree price discrimination:

$$CS_E^l(c_E) + mCS_E^s(c_E) \geq 2(CS_I^l(c_I) + mCS_I^s(c_I))$$

we see that

$$CS_I^l(p_I^{l,best}) + R_I^{l,best} + m(CS_I^s(p_I^{s,best}) + R_I^{s,best}) < 2(CS_I^l(c_I) + mCS_I^s(c_I))$$

and so entry is feasible even for values of $c_I < 1 - \sqrt{1/2}$, where it would not have been possible under third-degree price discrimination (see Proposition 9).

Case 2: The "first-best" solution is not feasible because the large buyer's self-selection condition is violated under this solution. Then, the "second-best" solution is:

$$\begin{aligned} p_E^{l*} &= c_E = 0 \\ p_E^{s*} &= \frac{mc_E + 1 - \frac{K}{m} \frac{1}{1-K}}{m + 1 - \frac{K}{m} \frac{1}{1-K}} > c_E = 0 \\ R_E^{s*} &= CS_I^s(p_I^{s,best}) + R_I^{s,best} - CS_E^s(p_E^{s*}) \\ R_E^{l*} &= CS^{l,net}(p_E^{s*}, q_E^s(p_E^{s*})) + R_E^{s*} - CS_E^l(c_E) \end{aligned}$$

provided that constraints (i), (v), and (vi) are satisfied under this solution. Constraints (ii) and (iii) are binding. We will now argue that constraint (iv) is implied to hold under the "second-best" solution as well: Inserting the solution values into constraint (iv), the inequality reads:

$$CS_E^s(p_E^{s*}) \geq CS^{s,net}(c_E, q_E^l(c_E)) + (CS^{l,net}(p_E^{s*}, q_E^s(p_E^{s*})) - CS_E^l(c_E))$$

Inserting for the consumer surplus terms from equations (2.6), (2.7), (2.8) and (2.9), the inequality reduces to

$$1 - K \geq \frac{K}{m} (1 - p_E^{s*})^2$$

which holds by the assumption that $1 - K > \frac{K}{m}$.

Case 3: The "second-best" solution is not feasible either because constraint (i) is violated under this solution. Then, constraints (i), (ii), and (iii) determine the "third-best" solution:

$$p_E^{ls} = c_E = 0$$

$$p_E^{ss} = 1 - \sqrt{2 \frac{m}{K} \frac{1-K}{1-K-\frac{K}{m}} \left(CS_I^l(p_I^{l,best}) + R_I^{l,best} - \left(CS_I^s(p_I^{s,best}) + R_I^{s,best} \right) \right)}$$

$$R_E^{ss} = CS_I^s(p_I^{s,best}) + R_I^{s,best} - CS_E^s(p_E^{ss})$$

$$R_E^{ls} = CS_I^l(p_I^{l,best}) + R_I^{l,best} - CS_E^l(c_E)$$

provided that constraints (v) and (vi) are satisfied under this solution. Since constraint (i) holds with strict equality, we can apply the same reasoning developed in Case 1 to argue that constraint (iv) is implied to be satisfied under the "third-best" solution as well. If the "third-best" solution violates either constraint (vi) or (v), then entry is not feasible. \square

Appendix B: The case of full discrimination among buyers

We will now show that if the incumbent can even discriminate among the small buyers, not just between them and the large buyer, the efficiency barrier to entry will be even higher than under third-degree price discrimination. Consider a candidate equilibrium where all buyers buy from the entrant. What can the incumbent do to break this candidate equilibrium?

(1) as before: I can offer the total surplus to the large buyer so as to entice him away

from the entrant.

$$w_I^s = w_I^l = c_I; F_I^s = CS_I^s(c_I), F_I^l = -mCS_I^s(c_I).$$

(2) I can offer the total surplus to a critical number of small buyers such that the entrant just fails to reach the minimum size: Denote the minimum number of small buyers the entrant needs by m_E . It is defined as $m_E \in \{1, \dots, m\}$ such that the demand generated by the large buyer plus m_E small buyers is sufficient to reach the minimum size

$$(1 - K) + m_E \frac{K}{m} \geq \bar{s}$$

while losing one small buyer would mean that the entrant cannot reach the minimum size

$$(1 - K) + (m_E - 1) \frac{K}{m} < \bar{s}$$

Then, if I can discriminate between the small buyers, I can concentrate the entire surplus on a number m_I of small buyers such that the number of buyers remaining with the entrant just falls short of m_E :

$$m - m_I < m_E \text{ and } m - (m_I - 1) \geq m_E$$

Suppose the entrant's offers to the small buyers differ in some way (that's possible, because we allow both firms to discriminate among the small buyers). Then, let us rank the small buyers from 1 to m by the total surplus they receive from the entrant, starting with the one who is worst off (wlog, we can set $w_E^s = w_E^l = c_E$):

$$CS_E^1(c_E) - F_E^1 \leq CS_E^2(c_E) - F_E^2 \leq \dots \leq CS_E^m(c_E) - F_E^m$$

Then, the incumbent will want to address the first m_I buyers, i.e. the ones with rank $j = \{1, \dots, m_I\}$, and make an offer to them that beats the entrant's offer to each of them. This will be possible iff the sum of total surpluses of the first m_I buyers is (strictly or weakly) smaller than the total surplus that the incumbent can offer, i.e. iff

$$\sum_{j=1}^{m_I} CS_E^j(c_E) - F_E^j < CS_I^l(c_I) + mCS_I^s(c_I) \quad (2.12)$$

Thus, to prevent such "poaching" of small buyers, the entrant must offer at least $CS_I^l(c_I) + mCS_I^s(c_I)$ to the first m_I buyers (the ones he treats "worst"). As it turns out (proof below), the cheapest way to achieve this is to offer the *same* total surplus to all the small buyers, namely

$$CS_E^j(c_E) - F_E^j = \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I)) \text{ for all } j = 1, \dots, m$$

In this way, *any* subgroup of small buyers of size m_I has total surplus $CS_I^l(c_I) + mCS_I^s(c_I)$, i.e. condition (2.12) just fails to hold for all possible subgroups of small buyers of size m_I .

But for the entrant to break even, it must be sufficiently efficient to be able to make such offers where neither the large buyer nor the small buyers are "assailable", i.e.

$$\begin{aligned} CS_E^l(c_E) + mCS_E^s(c_E) &\geq CS_I^l(c_I) + mCS_I^s(c_I) + m\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I)) \\ &= \left(1 + \frac{m}{m_I}\right) (CS_I^l(c_I) + mCS_I^s(c_I)) \end{aligned}$$

Since $m_I \in \{1, \dots, m\}$, we have that $1 + \frac{m}{m_I} \geq 2$, i.e. the entrant's feasibility constraint has tightened compared to the case where firms cannot discriminate among small buyers.

Claim: Suppose that I 's offer would have to be *strictly* better than E 's to induce buyers to switch. Let $m_I < m$. Consider all possible offers that E can make to the small buyers

such that the following condition is satisfied for all possible subgroups of small buyers of size m_I :

$$\sum_{j=1}^{m_I} CS_E^j(c_E) - F_E^j \geq CS_I^l(c_I) + mCS_I^s(c_I)$$

i.e. such that I cannot prevent E 's entry by poaching m_I small buyers. Then, the cheapest among all these offers is the one where

$$CS_E^j(c_E) - F_E^j = \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I)) \text{ for all } j = 1, \dots, m$$

Proof: Suppose instead that E makes some discriminatory offer. Rank all the small buyers by the total surplus they receive under E 's offer, starting with the buyer who is worst off. The offer we consider must satisfy

$$\sum_{j=1}^{m_I} CS_E^j(c_E) - F_E^j \geq CS_I^l(c_I) + mCS_I^s(c_I)$$

for the first m_I buyers. Since offers are discriminatory, there must be at least one buyer, call him j^* , whose total surplus is strictly below $\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$, i.e.

$$CS_E^{j^*}(c_E) - F_E^{j^*} < \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$$

(if every small buyer had at least $\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$, and some (or all) buyers had strictly more than that, this offer cannot be cheaper than the uniform offer where each buyer gets exactly $\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$). Since

$$\sum_{j=1}^{m_I} CS_E^j(c_E) - F_E^j \geq CS_I^l(c_I) + mCS_I^s(c_I),$$

buyer j^* must be in the group of the "poorest" small buyers, $j = \{1, \dots, m_I\}$, and at least one other buyer in this group (buyer m_I for sure, and maybe others) must have strictly more than $\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$, i.e.

$$CS_E^j(c_E) - F_E^j > \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I)) \text{ for some } j = \{2, \dots, m_I\}$$

(otherwise, their total surpluses could not sum up to $CS_I^l(c_I) + mCS_I^s(c_I)$). Moreover, all buyers with rank higher than m_I must have strictly more than $\frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$ as well (by the very nature of the ranking, all buyers $j = \{m_I + 1, \dots, m\}$ must have at least as much as buyer m_I , and by the assumption that $m_I < m$, there is at least one such buyer).

But then, there is at least one subgroup of small buyers of size m_I such that their total surpluses sum up to strictly more than $CS_I^l(c_I) + mCS_I^s(c_I)$: Consider the subgroup of the poorest buyers, $j = \{1, \dots, m_I\}$, and replace buyer 1 by buyer $m_I + 1$. We know that buyer 1 has total surplus $CS_E^1(c_E) - F_E^1 < \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$ (either $j^* = 1$, or $j^* > 1$, so that buyer 1 has weakly less total surplus than buyer j^*), while buyer $m_I + 1$ has total surplus $CS_E^{m_I+1}(c_E) - F_E^{m_I+1} > \frac{1}{m_I} (CS_I^l(c_I) + mCS_I^s(c_I))$. Thus, we have that

$$(CS_E^{m_I+1}(c_E) - F_E^{m_I+1}) - (CS_E^1(c_E) - F_E^1) = \Delta > 0$$

and so the sum of total surpluses of group $j = \{2, \dots, m_I + 1\}$ is

$$\begin{aligned} & \left(\sum_{j=1}^{m_I} CS_E^j(c_E) - F_E^j \right) - (CS_E^1(c_E) - F_E^1) + (CS_E^{m_I+1}(c_E) - F_E^{m_I+1}) \\ & \geq CS_I^l(c_I) + mCS_I^s(c_I) + \Delta \\ & > CS_I^l(c_I) + mCS_I^s(c_I) \end{aligned}$$

Thus, the discriminatory offer under consideration must be strictly more expensive for E than the uniform offer

$$CS_E^j(c_E) - F_E^j = \frac{1}{m_I} (CS_I^l(c_I) + m CS_I^s(c_I)) \text{ for all } j = 1, \dots, m$$

where no subgroup of size m_I can ever have strictly more total surplus than $CS_I^l(c_I) + m CS_I^s(c_I)$. \square

Appendix C: Uniform two-part tariffs

Let firms $i = I, E$ use tariffs of the following form:

$$T_i = p_i q_i^j - R_i$$

i.e. firms cannot discriminate among buyers, neither by type nor by the quantity they buy, but they can charge two-part tariffs.

Miscoordination Equilibria

Case 1: I sells to all buyers

Suppose buyers are miscoordinated (buy from I whenever I offers non-negative surplus), and the incumbent wants to sell to all of them (not just to the large buyer). Then, the incumbent's problem reads:

$$\max_{\{p_I, R_I\}} (p_I - c_I) (1 - p_I) - (m + 1) R_I$$

subject to the small and large buyer's participation constraints (PC):

$$(i) \quad \frac{1}{2} \frac{K}{m} (1 - p_I)^2 + R_I \geq 0$$

$$(ii) \quad \frac{1}{2} (1 - K) (1 - p_I)^2 + R_I \geq 0$$

Note that $1 - K > \frac{K}{m}$ implies that the large buyer's PC will always be satisfied when the small buyer's PC holds, so that only constraint (i) can be binding.

Hence, at any optimum where I sells to all buyers, we must have

$$R_I^* = -\frac{1}{2} \frac{K}{m} (1 - p_I^*)^2$$

Inserting into the objective function and solving for the optimal unit price, we obtain

$$p_I^* = \frac{1 + c_I - \frac{m+1}{m} K}{2 - \frac{m+1}{m} K}$$

Note that $p_I^* > c_I$, i.e. the incumbent will charge a price above marginal cost. This implies efficiency losses, and reduces the rent that I can extract from the small buyers. But these losses are more than compensated by the profits I makes on the sales to the large buyer, where I earns a positive markup on each unit sold (in particular on the units exceeding $q_I^s(p_I)$).

Case 2: I sells to the large buyer only

The incumbent may find it more profitable to sell to the large buyer only. In this case, I will optimally set

$$p_I = c_I; R_I = -\frac{1}{2} (1 - K) (1 - c_I)^2$$

Then, the incumbent will fully extract the large buyer's surplus, while the small buyers will not want to buy at all (because that would leave them with negative surplus).

When will it be optimal to sell to the large buyer only, rather than to all buyers?

$$\frac{1}{2} (1 - K) (1 - c_I)^2 \geq (p_I^* - c_I) (1 - p_I^*) - (m + 1) R_I^*$$

This inequality can be reduced to

$$K \leq \left(\frac{m}{m+1} + \frac{1}{2} \right) - \sqrt{\left(\frac{m}{m+1} \right)^2 + \left(\frac{1}{2} \right)^2} = K^* < 0.5$$

Thus, it will be worthwhile to sell to the large buyer only whenever the latter is fairly large compared to the group of small buyers.

Entry Equilibria

Following our reasoning developed for the case of discriminatory two-part tariffs, the incumbent can break entry by concentrating rent on the small buyers (thus forcing the large buyer to buy from I as well) or vice versa. Let us start with the best offer I can make to the small buyers.

I 's best offer to the small buyers

I will have to solve the following problem:

$$\max_{\{p_I, R_I\}} \frac{1}{2} \frac{K}{m} (1 - p_I)^2 + R_I$$

subject to the break-even constraint and the large buyer's participation constraint (PC):

$$(i) \quad (p_I - c_I) (1 - p_I) - (m + 1) R_I \geq 0$$

$$(ii) \quad \frac{1}{2} (1 - K) (1 - p_I)^2 + R_I \geq 0$$

Note that the large buyer's PC will never be binding: if $\frac{1}{2} \frac{K}{m} (1 - p_I)^2 + R_I \geq 0$ (which will certainly hold at the optimum), then $1 - K > \frac{K}{m}$ implies that the large buyer's PC must

be satisfied as well. The break-even constraint will hold with strict equality, which yields

$$R_I^{s*} = \frac{1}{m+1} (p_I^{s*} - c_I) (1 - p_I^{s*})$$

Then, our problem solves for the optimal unit price as

$$p_I^{s*} = \frac{1 + c_I - \frac{m+1}{m} K}{2 - \frac{m+1}{m} K}$$

which is exactly the same p_I as in Case 1 of Miscoordination.

To the typical small buyer, this offer gives total surplus of

$$CS_I^s(p_I^{s*}) + R_I^{s*} = \left[\frac{1 - c_I}{2 - \frac{m+1}{m} K} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2} \frac{K}{m} \right]$$

I 's best offer to the large buyer

I 's best offer to the large buyer will depend on parameter values. I has to solve

$$\max_{\{p_I, R_I\}} \frac{1}{2} (1 - K) (1 - p_I)^2 + R_I$$

subject to the break-even constraint and the small buyers' participation constraint (PC):

$$(i) \quad (p_I - c_I) (1 - p_I) - (m+1) R_I \geq 0$$

$$(ii) \quad \frac{1}{2} \frac{K}{m} (1 - p_I)^2 + R_I \geq 0$$

$$\text{Case 1: } K \geq \max \left\{ \frac{m - c_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$$

Then, the break-even constraint will be binding at the optimum, while the small buyers'

PC will be oversatisfied. We can solve the problem analogously to the one before to obtain

Solution II:

$$p_I^{I*} = \frac{1 + c_I - (m+1)(1-K)}{2 - (m+1)(1-K)} \text{ and } R_I^{I*} = \frac{1}{m+1} (p_I^{I*} - c_I) (1 - p_I^{I*}) \quad (2.13)$$

The second-order condition requires that $2 - (m + 1)(1 - K) > 0$, which is equivalent to

$$K > \frac{m - 1}{m + 1}$$

If $K \geq \max \left\{ \frac{m - c_I}{m + 1}, \frac{m}{m + 1} \frac{2m}{2m + 1} \right\}$, then the SOC is always satisfied.

Now, for p_I^{l*} to be non-negative, i.e. $p_I^{l*} \geq 0$, we must have

$$K \geq \frac{m - c_I}{m + 1}$$

which is again implied by our lower bounds on K . Hence, $p_I^{l*} \in (0, c_I)$.

Finally, for the small buyers' PC to hold under this solution, we can rearrange constraint (ii) to obtain

$$K \geq \frac{m}{m + 1} \frac{2m}{2m + 1}$$

which will of course hold under the lower bounds assumed for K .

Note that $\max \left\{ \frac{m - c_I}{m + 1}, \frac{m}{m + 1} \frac{2m}{2m + 1} \right\} = \frac{m - c_I}{m + 1}$ whenever $K \geq \frac{2mc_I}{m + 1}$.

Case 2: $\frac{2mc_I}{m + 1} \leq K < \frac{m - c_I}{m + 1}$

If $K < \max \left\{ \frac{m - c_I}{m + 1}, \frac{m}{m + 1} \frac{2m}{2m + 1} \right\}$, this means that solution I1 either violates the small buyers' PC, or the non-negativity of p_I^{l*} . Let $K \geq \frac{2mc_I}{m + 1}$, so that the non-negativity constraint is stronger than the small buyers' PC. Then, the best offer that I can make to the large buyer is one where $p_I^{l*} = 0$ and the small buyers' PC (ii) is binding, provided that this is feasible, i.e. the break-even constraint (i) is not violated. Solution I2 is characterized by

$$p_I^{l*} = 0 \text{ and } R_I^{l*} = -\frac{1}{2} \frac{K}{m} \quad (2.14)$$

Solution I2 trivially satisfies both the small buyers' PC and the non-negativity of p_I^I .

For the incumbent to break even under this solution, we must have

$$(0 - c_I) + (m + 1) \frac{1}{2} \frac{K}{m} \geq 0$$

which reduces to

$$K \geq \frac{2mc_I}{m+1}$$

This condition corresponds to the lower bound on K imposed for Case 2. Note that solution I1 always yields higher payoff for the large buyer than solution I2; thus, if K is sufficiently large to admit both solutions, I will always choose I1.

Case 3: $K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$

If $K < \frac{2mc_I}{m+1}$, then $\max \left\{ \frac{m-c_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\} = \frac{m}{m+1} \frac{2m}{2m+1}$. Thus, $K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$

implies that

(i) solution I1 would violate the small buyers' PC (and may or may not satisfy the non-negativity constraint on p_I^I), and

(ii) solution I2 would imply losses for I .

Hence, the only feasible solution is the one where both the small buyers' PC and the break-even constraint are binding. These two constraints will then fully determine p_I^I and R_I^I of **Solution I3** as:

$$p_I^I = \frac{2c_I - (m+1) \frac{K}{m}}{2 - (m+1) \frac{K}{m}} \text{ and } R_I^I = -\frac{1}{2} \frac{K}{m} \left[\frac{1 - c_I}{1 - \frac{m+1}{2m} K} \right]^2 \quad (2.15)$$

Note that our parameter restriction, $K < \frac{2mc_I}{m+1}$, implies that $p_I^I \in (0, c_I)$, and that $\left[\frac{1 - c_I}{1 - \frac{m+1}{2m} K} \right]^2 <$

1 (so that the large buyer's total surplus under solution I3 is smaller than under solution I2).

The following table summarizes the maximum total payoff that I can offer to the large buyer, $CS_I^l(p_I^{l*}) + R_I^{l*}$:

I's best offer to large buyer		
sol'n	applies if	LB's payoff
I1	$K \geq \max \left\{ \frac{m-c_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$	$\left[\frac{1-c_I}{2-(m+1)(1-K)} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2}(1-K) \right]$
I2	$\frac{2mc_I}{m+1} \leq K < \frac{m-c_I}{m+1}$	$\frac{1}{2} \left(1 - K - \frac{K}{m} \right)$
I3	$K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$	$\frac{1}{2} \left(1 - K - \frac{K}{m} \right) \left[\frac{1-c_I}{1-\frac{m+1}{2m}K} \right]^2$

E 's best offer

Recall that for a pure-strategy entry equilibrium to exist, the following conditions must be satisfied:

- the entrant matches both I 's best offer to the small buyers and I 's best offer to the large buyer;
- the entrant breaks even; and
- the entrant generates enough demand to reach the minimum size.¹²

Thus, E 's problem is:

$$\max_{\{p_E, R_E\}} (p_E - c_E)(1 - p_E) - (m+1)R_E$$

subject to:

¹²This implies, of course, that E must serve *all* buyers, i.e. serving only the large buyer will never be an option for E (even if it is for I under a miscoordination equilibrium).

- (i) $\frac{1}{2} \frac{K}{m} (1 - p_E)^2 + R_E \geq CS_I^s(p_I^{s*}) + R_I^{s*}$
- (ii) $\frac{1}{2} (1 - K) (1 - p_E)^2 + R_E \geq CS_I^l(p_I^{l*}) + R_I^{l*}$
- (iii) $(p_E - c_E) (1 - p_E) - (m + 1) R_E \geq 0$
- (iv) $1 - p_E \geq \bar{s}$

Let us first study under which conditions entry is possible if condition (iv) is not binding, i.e. the minimum size is low enough for the entrant to have some degrees of freedom in setting p_E .

In this case, at least one of the two participation constraints, (i) and (ii), must be binding. As it turns out, the small buyers' PC will always be binding, while the large buyer's IC may or may not be binding.¹³

Case 1: Consider a solution where the small buyers' participation constraint (i) is binding, while the large buyer's PC (ii) is oversatisfied. Then, (i) determines E 's fixed payment as

$$R_E^* = CS_I^s(p_I^{s*}) + R_I^{s*} - \frac{1}{2} \frac{K}{m} (1 - p_E^*)^2$$

Inserting into the objective function and solving for the optimal p_E , we obtain **Solution E1**

$$p_E^* = \frac{1 + c_E - \frac{m+1}{m} K}{2 - \frac{m+1}{m} K} \quad (2.16)$$

$$R_E^* = \left[\frac{1 - c_I}{2 - \frac{m+1}{m} K} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2} \frac{K}{m} \right] - \frac{1}{2} \frac{K}{m} \left(\frac{1}{2 - \frac{m+1}{m} K} \right)^2 \quad (2.17)$$

which is analogous to I 's optimal offer under the Case 1 miscoordination equilibrium.

¹³In other words, there is no solution where (ii) is binding and (i) is oversatisfied.

The optimal unit price will generate enough demand if

$$1 - p_E^* = \frac{1}{2 - \frac{m+1}{m}K} \geq \bar{s} > \max \{K, 1 - K\}$$

A necessary condition for this inequality to hold is that $1 - p_E^* > \max \{K, 1 - K\}$, which can be rearranged to have

$$K > \left(\frac{m}{m+1} + \frac{1}{2} \right) - \sqrt{\left(\frac{m}{m+1} \right)^2 + \left(\frac{1}{2} \right)^2} = K^*$$

This condition corresponds exactly to the condition under which the incumbent will sell to all buyers (rather than just to the large buyer) under the miscoordination equilibria discussed in Section 2.

We still have to verify whether the large buyer's PC (ii) holds under this solution:

(i) Let $K \geq \max \left\{ \frac{m-c_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$, so that

$$CS_I^l(p_I^{l*}) + R_I^{l*} = \left[\frac{1 - c_I}{2 - (m+1)(1-K)} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2}(1-K) \right]$$

Then, $\frac{1}{2}(1-K)(1-p_E)^2 + R_E \geq CS_I^l(p_I^{l*}) + R_I^{l*}$ reduces to

$$2 - (m+1)(1-K) \geq (1-c_I)^2 \left(2 - \frac{m+1}{m}K \right)$$

Since $K \geq \frac{m-c_I}{m+1}$ whenever solution I1 applies, this inequality can be simplified to

$$1 \geq (1-c_I) \left(1 + \frac{c_I}{m} \right)$$

which is always true.

(ii) Let $\frac{2mc_I}{m+1} \leq K < \frac{m-c_I}{m+1}$, so that solution I2 applies, and hence

$$CS_I^l(p_I^{l*}) + R_I^{l*} = \frac{1}{2} \left(1 - K - \frac{K}{m} \right)$$

Then, $\frac{1}{2}(1-K)(1-p_E)^2 + R_E \geq CS_I^l(p_I^{l*}) + R_I^{l*}$ reduces to

$$(1-c_I)^2 \geq \frac{(1-K-\frac{K}{m})\left((2-\frac{m+1}{m}K)^2-1\right)}{\frac{1}{m+1}\left(2-\frac{m+1}{m}K\right)} \quad (2.18)$$

This inequality may or may not be satisfied (it will fail to hold if $c_I \rightarrow 0$ and $K \rightarrow 0$).

(iii) Let $K < \min\left\{\frac{2mc_I}{m+1}, \frac{m}{m+1}\frac{2m}{2m+1}\right\}$, so that solution I3 applies, and hence

$$CS_I^l(p_I^{l*}) + R_I^{l*} = \frac{1}{2}\left(1-K-\frac{K}{m}\right)\left[\frac{1-c_I}{1-\frac{m+1}{2m}K}\right]^2$$

Then, $\frac{1}{2}(1-K)(1-p_E)^2 + R_E \geq CS_I^l(p_I^{l*}) + R_I^{l*}$ reduces to

$$(1-c_I)^2 \leq \frac{\frac{1}{2}\left(1-K-\frac{K}{m}\right)}{2\left(1-K-\frac{K}{m}\right)-\frac{1}{m+1}+\frac{1}{2}\frac{K}{m}} \quad (2.19)$$

Again, this inequality may or may not be satisfied (it will fail to hold if $c_I \rightarrow 0$ and $K \rightarrow 0$).

Recall that if $1-p_E^* > \max\{K, 1-K\}$ is violated, then p_E^* cannot possibly generate enough demand for the entrant to reach the minimum size. Now, if either I1 or I2 applies, and solution E1 satisfies the large buyer's PC, it will also satisfy $1-p_E^* > \max\{K, 1-K\}$. If instead I3 applies, then $1-p_E^* > \max\{K, 1-K\}$ is not implied by the large buyer's PC, but constitutes an additional constraint on the applicability of solution E1.

Case 2: Suppose the large buyer's PC fails to hold under solution E1, i.e. either I2 applies and condition (2.18) is violated, or I3 applies and condition (2.19) is violated. Then, both the large and small buyers' PC must be binding, which will fully determine p_E and R_E of **Solution E2:**

$$\begin{aligned} p_E^* &= 1 - \sqrt{2 \frac{(CS_I^l(p_I^{l*}) + R_I^{l*}) - (CS_I^s(p_I^{s*}) + R_I^{s*})}{1-K-\frac{K}{m}}} \\ R_E^* &= (CS_I^s(p_I^{s*}) + R_I^{s*}) \frac{1-K}{1-K-\frac{K}{m}} - (CS_I^l(p_I^{l*}) + R_I^{l*}) \frac{\frac{K}{m}}{1-K-\frac{K}{m}} \end{aligned} \quad (2.20)$$

where $CS_I^l(p_I^{l*}) + R_I^{l*}$ depends on whether I2 or I3 applies.

Of course, solution E1 will be more profitable than E2, so that the entrant will want to apply E1 whenever possible, and will apply E2 only if necessary.

Note that $p_E^* > 0$ is equivalent to

$$\frac{1}{2} \left(1 - K - \frac{K}{m} \right) > (CS_I^l(p_I^{l*}) + R_I^{l*}) - (CS_I^s(p_I^{s*}) + R_I^{s*})$$

which is always satisfied because $CS_I^l(p_I^{l*}) + R_I^{l*} = \frac{1}{2} (1 - K - \frac{K}{m})$ under I2 and $CS_I^l(p_I^{l*}) + R_I^{l*} < \frac{1}{2} (1 - K - \frac{K}{m})$ under I3, while $CS_I^s(p_I^{s*}) + R_I^{s*} > 0$.

Again, this solution can only apply if p_E^* generates enough demand for the entrant to reach the minimum size, i.e. if

$$1 - p_E^* \geq \bar{s} > \max \{K, 1 - K\}$$

(we will come back to this later).

Will the entrant break even?

After determining the incumbent's best offers to both large and small buyers, as well as the entrant's best offer when E has to match I 's best offers, let us now turn to the question whether the entrant will break even under these offers.

Suppose solution E1 applies. Then, inserting

$$p_E^* = \frac{1 + c_E - \frac{m+1}{m}K}{2 - \frac{m+1}{m}K} \quad (2.21)$$

$$R_E^* = \left[\frac{1 - c_I}{2 - \frac{m+1}{m}K} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2m}K \right] - \frac{1}{2m}K \left(\frac{1}{2 - \frac{m+1}{m}K} \right)^2 \quad (2.22)$$

into the entrant's break-even constraint,

$$(p_E - c_E)(1 - p_E) - (m+1)R_E \geq 0$$

we can reduce this inequality to

$$(1 - c_I)^2 < 1$$

which is always satisfied.

Thus, we can conclude that if E1 applies, the entrant will always break even, no matter what I 's best offer to the large buyer is (i.e. no matter if I1, I2, or I3 applies).

Next, suppose that the entrant has to apply E2. Then, we have to distinguish the following cases:

(i) Let $\frac{2mc_I}{m+1} \leq K < \frac{m-c_I}{m+1}$, so that solution I2 applies:

$$CS_I^l(p_I^*) + R_I^* = \frac{1}{2} \left(1 - K - \frac{K}{m} \right)$$

Then, the entrant's break-even constraint can be simplified to

$$\begin{aligned} & \left(1 - K - \frac{K}{m} \right)^2 \left(1 - \frac{1}{4} \left(2 - \frac{m+1}{m} K \right)^2 \right) - (1 - c_I)^4 \frac{\left(\frac{1}{m+1} - \frac{1}{2} (1 - K) \right)^2}{\left(2 - \frac{m+1}{m} K \right)^2} \geq \\ & \geq \left(1 - K - \frac{K}{m} \right) (1 - c_I)^2 \frac{1}{m+1} \left(\frac{1 - \left(1 - \frac{1}{2} (m+1) (1 - K) \right) \left(2 - K \frac{m+1}{m} \right)}{\left(2 - K \frac{m+1}{m} \right)} \right) \end{aligned}$$

As it turns out, the condition

$$1 - p_E^* > \max \{K, 1 - K\}$$

(which is necessary for E to reach the minimum size) is satisfied whenever the break-even constraint holds.

(ii) Let $K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1}, \frac{2m}{2m+1} \right\}$, so that solution I3 applies:

$$CS_I^l(p_I^*) + R_I^* = \frac{1}{2} \left(1 - K - \frac{K}{m} \right) \left[\frac{1 - c_I}{1 - \frac{m+1}{2m} K} \right]^2$$

Then, the entrant's break-even constraint can be simplified to

$$(1 - c_I)^2 \leq \frac{2 \left(1 - K - \frac{K}{m}\right) \left(2 \left(1 - K - \frac{K}{m}\right) - \frac{1}{m+1} + \frac{1}{2} \frac{K}{m}\right)}{\left(2 \left(1 - K - \frac{K}{m}\right) - \frac{1}{m+1} + \frac{1}{2} (1 - K)\right)^2}$$

Recall that if $1 - p_E^* > \max \{K, 1 - K\}$ is violated, then p_E^* cannot possibly generate enough demand for the entrant to reach the minimum size. We argued that if I2 applies, and solution E2 satisfies the break-even constraint, it will also satisfy $1 - p_E^* > \max \{K, 1 - K\}$. If instead I3 applies, then $1 - p_E^* > \max \{K, 1 - K\}$ is not implied by the break-even constraint, but constitutes an additional constraint on the applicability of solution E2. Note that this constraint can only be binding for values of K below K^* .

What if constraint (iv) is binding?

Case 1: Let the minimum size be as low as possible, i.e. $\bar{s} = \max \{K, 1 - K\} + \varepsilon$. Then, we argued above that

(i) under solution E1, the entrant's optimal unit price will violate $1 - p_E^* \geq \bar{s}$ whenever $1 - p_E^* \leq \max \{K, 1 - K\}$, which we could simplify to $K \leq K^*$;

(ii) under solution E2 and I2, the optimal unit price will always satisfy $1 - p_E^* \geq \bar{s}$;

(iii) under solution E2 and I3, condition $1 - p_E^* \geq \bar{s}$ represents an additional constraint, which can only be binding for values of K below K^* .

Thus, in either case, $1 - p_E^* \geq \bar{s}$ can only be binding when $K < 0.5$, so that $\max \{K, 1 - K\} = 1 - K$, and therefore

$$\bar{s} = 1 - K + \varepsilon$$

Now, let the entrant charge

$$p_E^* = 1 - \bar{s} = K - \varepsilon$$

i.e. the highest possible price that generates enough demand for E to reach the minimum size. Then, the entrant will break even if its rebate payments satisfy:

$$R_E^* \leq \frac{1}{m+1} (K - \varepsilon) (1 - K + \varepsilon)$$

Under this offer, the small and large buyer's PC will be satisfied if

$$\begin{aligned} \frac{1}{2} \frac{K}{m} (1 - K + \varepsilon)^2 + R_E^* &\geq \left[\frac{1 - c_I}{2 - \frac{m+1}{m} K} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2} \frac{K}{m} \right] \\ \frac{1}{2} (1 - K) (1 - K + \varepsilon)^2 + R_E^* &\geq \frac{1}{2} \left(1 - K - \frac{K}{m} \right) \left[\frac{1 - c_I}{1 - \frac{m+1}{2m} K} \right]^2 \end{aligned}$$

Where solution E3 violates the break-even constraint, the two PCs can never hold simultaneously. Where solution E3 violates $1 - p_E^* \geq \bar{s}$ but satisfies the break-even constraint, the large buyer's PC will always hold under $p_E^* = 1 - \bar{s}$. Thus, it will be the small buyer's PC that determines feasibility of our candidate solution.

Case 2: If $\bar{s} = 1$, the entrant has no choice but to set $p_E = 0$ (marginal cost pricing) to generate enough demand: $1 - p_E = 1 = \bar{s}$. The break-even constraint implies that the entrant cannot make any positive payments to the buyers, i.e. $R_E \leq 0$.

Then, entry will be feasible if:

$$\begin{aligned} \frac{1}{2} \frac{K}{m} &\geq CS_I^s(p_I^{s*}) + R_I^{s*} \text{ and} \\ \frac{1}{2} (1 - K) &\geq CS_I^l(p_I^{l*}) + R_I^{l*} \end{aligned}$$

If solution I1 applies, then both the large and small buyers' PC will be satisfied under ($p_E = R_E = 0$); thus, if $K \geq \max \left\{ \frac{m-c_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$, entry is always feasible.

If either solution I2 or I3 applies, the large buyer's PC is always satisfied under ($p_E = R_E = 0$).

$$CS_I^l(p_I^{l*}) + R_I^{l*} \leq \frac{1}{2} \left(1 - K - \frac{K}{m} \right) < \frac{1}{2} (1 - K)$$

so that the small buyers' PC will determine whether entry is feasible or not.

Conclusion

We can conclude that entry will never be feasible in the following cases:

(i) If $\frac{2mc_I}{m+1} \leq K < \frac{m-c_I}{m+1}$, so that solution I2 applies, and the entrant cannot play solution

E1 because condition (2.18) is violated:

$$(1 - c_I)^2 < \frac{\left(1 - K - \frac{K}{m}\right) \left(\left(2 - \frac{m+1}{m} K\right)^2 - 1 \right)}{\frac{1}{m+1} \left(2 - \frac{m+1}{m} K\right)}$$

nor can the entrant play E2, because the break-even constraint fails to hold

$$\begin{aligned} & \left(1 - K - \frac{K}{m}\right)^2 \left(1 - \frac{1}{4} \left(2 - \frac{m+1}{m} K\right)^2\right) - (1 - c_I)^4 \frac{\left(\frac{1}{m+1} - \frac{1}{2} (1 - K)\right)^2}{\left(2 - \frac{m+1}{m} K\right)^2} < \\ & < \left(1 - K - \frac{K}{m}\right) (1 - c_I)^2 \frac{1}{m+1} \left(\frac{1 - \left(1 - \frac{1}{2} (m+1) (1 - K)\right) \left(2 - K \frac{m+1}{m}\right)}{\left(2 - K \frac{m+1}{m}\right)} \right) \end{aligned}$$

or

(ii) $K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$, so that solution I3 applies, and the entrant cannot play

solution E1 because condition (2.19) is violated

$$(1 - c_I)^2 > \frac{\frac{1}{2} \left(1 - K - \frac{K}{m}\right)}{2 \left(1 - K - \frac{K}{m}\right) - \frac{1}{m+1} + \frac{1}{2} \frac{K}{m}}$$

nor can the entrant play E2, because the break-even constraint fails to hold

$$(1 - c_I)^2 > \frac{2 \left(1 - K - \frac{K}{m}\right) \left(2 \left(1 - K - \frac{K}{m}\right) - \frac{1}{m+1} + \frac{1}{2} \frac{K}{m}\right)}{\left(2 \left(1 - K - \frac{K}{m}\right) - \frac{1}{m+1} + \frac{1}{2} (1 - K)\right)^2}$$

or

(iii) $K < \min \left\{ \frac{2mc_I}{m+1}, \frac{m}{m+1} \frac{2m}{2m+1} \right\}$, so that solution I3 applies, and the entrant cannot play either solution E1 or E2 because constraint (iv) is binding ($1 - p_E^* < \bar{s} = 1 - K + \varepsilon$), and the offer which satisfies constraint (iv)

$$p_E^* = 1 - \bar{s} = K - \varepsilon$$

$$R_E^* \leq \frac{1}{m+1} (K - \varepsilon) (1 - K + \varepsilon)$$

does not satisfy the small buyer's PC:

$$\frac{1}{2} \frac{K}{m} (1 - K + \varepsilon)^2 + R_E^* < \left[\frac{1 - c_I}{2 - \frac{m+1}{m} K} \right]^2 \left[\frac{1}{m+1} - \frac{1}{2} \frac{K}{m} \right]$$

Figure 2.2 illustrates this region in the (c_I, K) plane, with $(1 - c_I)^2$ on the x-axis and K on the y-axis. Above the black diagonal line (where $K \geq \frac{2mc_I}{m+1}$), Solution I2 applies, while below that line, I will play solution I3. Solution E2 applies below the other black line, where solution E1 is no longer feasible because the large buyer's PC is violated. The black dotted line represents $(1 - c_I)^2 = \frac{1}{2}$; to the right of this line, there are no entry equilibria under third-degree price discrimination. Entry will never be possible in the lower right-hand corner, everywhere below the red line. Entry is always feasible above the green line, and may be feasible in the region between the green and red line provided the minimum threshold \bar{s} is low enough.

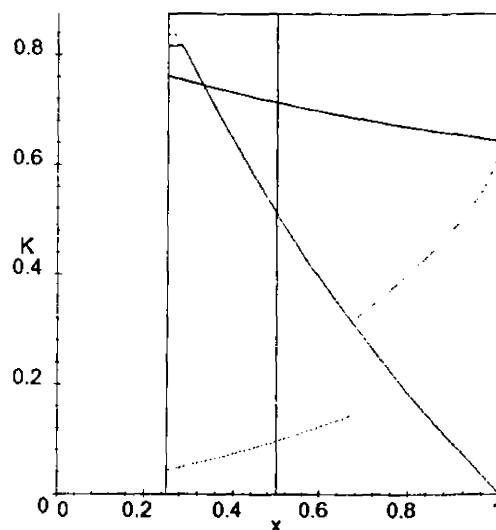


Figure 2.2: Entry will never be possible below the red line, and is always possible above the green line.

Note that a large part of the parameter space where entry would be impossible under third-degree price discrimination does belong to the region where entry is (or may be) feasible under uniform two-part tariffs (this is the region to the right of the dotted line and above the red line). On the other hand, there is a small region to the left of the dotted line where entry cannot occur under uniform two-part tariffs (namely below the red line), but would be possible under third-degree price discrimination.

In other words, there is no clear-cut answer as to which pricing regime is more exclusionary than the other: if the incumbent is rather efficient and the large buyer is relatively small, discriminatory tariffs can be exclusionary while uniform tariffs are not. On the other hand, if the incumbent is very inefficient and the large buyer is very large, discriminatory

tariffs will allow for entry when uniform tariffs do not.

Welfare Analysis

The following table shows the large and small buyers' surplus, firms' profits, and welfare loss under all possible scenarios discussed above.

Compared to the cases of third- and second-degree discrimination, the most remarkable finding is that full efficiency cannot even be obtained when the entrant serves, because the entrant's price will always be $p_E^* > c_E = 0$ (with the exception of the very special case of $\bar{s} = 1$). The reason is that price equal to marginal cost can never be profit-maximizing if the firm must charge uniform tariffs: it always pays to increase the price by an ε , because a positive mark-up allows the firm to appropriate more of the large buyer's rent, which will more than compensate the loss on the small buyers.

Thus, the main conclusion is that to achieve full efficiency, some discrimination must be possible (recall that second-degree price discrimination was already sufficient, no need for third-degree PD). This result is not surprising in the light of the well-known welfare effects of price discrimination in general.

Moreover, note that the buyers' surplus when E serves is determined by the best offer that I can make to the buyers. Thus, the better the offers that the incumbent can make (and these offers will be the better, the more I can discriminate), the more rent will end up in the hands of buyers when E serves. In other words, even if the standard of comparison is buyer surplus (rather than total welfare), there is a case to be made for discrimination.

Table 1: Welfare under Uniform Two-Part Tariffs

(a) Buyers' Net Total Surplus

	Small Buyers	Large Buyer
<i>I</i> serves LB	0	0
<i>I</i> serves all	0	$\frac{1}{2} \left(1 - K - \frac{K}{m}\right) \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2$
<i>E</i> serves (E1)	$\left(\frac{1}{m+1} - \frac{K}{2m}\right) \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2$	$\frac{\frac{1}{2} \left(1 - K - \frac{K}{m}\right) + \left(\frac{1}{m+1} - \frac{K}{2m}\right) (1-c_I)^2}{\left(2-\frac{m+1}{m}K\right)^2}$
<i>E</i> serves (E2, I2)	$\left(\frac{1}{m+1} - \frac{K}{2m}\right) \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2$	$\frac{1}{2} \left(1 - K - \frac{K}{m}\right)$
<i>E</i> serves (E2, I3)	$\left(\frac{1}{m+1} - \frac{K}{2m}\right) \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2$	$\frac{1}{2} \left(1 - K - \frac{K}{m}\right) \left(\frac{1-c_I}{1-\frac{m+1}{2m}K}\right)^2$

(b) Firms' Profits

	Incumbent	Entrant
<i>I</i> serves LB	$\frac{1}{2} (1 - K) (1 - c_I)^2$	0
<i>I</i> serves all	$\left(1 - \frac{m+1}{2m}K\right) \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2$	0
<i>E</i> serves (E1)	0	$\frac{c_I(2-c_I)\left(1-\frac{m+1}{2m}K\right)}{\left(2-\frac{m+1}{m}K\right)^2}$
<i>E</i> serves (E2, I2)	0	$\in \left[0, \frac{c_I(2-c_I)\left(1-\frac{m+1}{2m}K\right)}{\left(2-\frac{m+1}{m}K\right)^2}\right)$
<i>E</i> serves (E2, I3)	0	$\in \left[0, \frac{c_I(2-c_I)\left(1-\frac{m+1}{2m}K\right)}{\left(2-\frac{m+1}{m}K\right)^2}\right)$

(c) Welfare Loss

	Allocative	Productive	Total
<i>I</i> serves LB	$-\frac{K}{2} - \frac{1}{2} (1 - K) c_I^2$	$-c_I (1 - K) (1 - c_I)$	$\frac{1}{2} (1 - K) (1 - c_I)^2$
<i>I</i> serves all	$-\frac{1}{2} \left(\frac{1+c_I-\frac{m+1}{m}K}{2-\frac{m+1}{m}K}\right)^2$	$-c_I \left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)$	$\left(\frac{1-c_I}{2-\frac{m+1}{m}K}\right)^2 \left(\frac{3}{2} - K - \frac{K}{m}\right)$
<i>E</i> serves (E1)	$-\frac{1}{2} \left(\frac{1-\frac{m+1}{m}K}{2-\frac{m+1}{m}K}\right)^2$	0	$-\frac{1}{2} \left(\frac{1-\frac{m+1}{m}K}{2-\frac{m+1}{m}K}\right)^2$
<i>E</i> serves (E2, I2)	$= -\frac{1}{2} (p_E^* I2)^2$	0	$\in \left(-\frac{1}{2} \left(\frac{1-\frac{m+1}{m}K}{2-\frac{m+1}{m}K}\right)^2, 0\right)$
<i>E</i> serves (E2, I3)	$= -\frac{1}{2} (p_E^* I3)^2$	0	$\in \left(-\frac{1}{2} \left(\frac{1-\frac{m+1}{m}K}{2-\frac{m+1}{m}K}\right)^2, 0\right)$



Chapter 3

The Underground Economy Across Countries: Evading Taxes, or Evading Competition?

Introduction

The underground (or shadow) economy is broadly defined as economic activities which are concealed from public authorities to avoid the payment of taxes and social security contributions, and to avoid compliance with certain legal standards (e.g. labor market regulations, trade licenses). Unreported activities are a universal feature of economic life, and assume considerable proportions even in the industrialized world, where they are estimated to range between 8 and as much as 28 percent of official GDP.¹

It has been observed that the size of the underground economy (as a fraction of overall economic activity) varies considerably across countries, which has motivated an extensive literature investigating the causes of this particular form of regulation failure: The burden

¹See Schneider/Enste (2000) for estimates of the size of the underground economy in numerous countries and a critical discussion of the various measurement methods.

of taxes and social security contributions, excessive market regulation, as well as ineffective law enforcement and corruption, have been discussed at length to explain cross-country variations (see Schneider/Enste (2000), Johnson et al (1998), Lemieux et al (1994)).

While acknowledging that all of these factors do play an important role in determining the size of the underground economy, there are good reasons to believe that this list is not exhaustive. In particular, these factors do not explain the substantial variations in the share of the underground economy that have been observed not only *across* countries, but even *within* a single country, i.e. within the same legal and institutional framework. A well-documented example is the South of Italy, where the share of the underground economy is twice the national average (De Rita/Camusi (2003)).²

To explain such variations across regions within the same country, it has been argued that regions differ in the sectoral composition of their local industries, and that some sectors (e.g. construction, retail trade, household services) offer better opportunities to evade taxes than others (Kesselmann (1989)). Yet, empirical evidence suggests that even within the same sector, shadow-economic activity in the South of Italy is more pervasive than in the North (see Scarlato (2001)). Thus, the industry-composition argument cannot fully account for the inter-regional pattern observed in Italy.

In this chapter, I present a novel rationale for the variations in the share of the underground economy which can explain both inter-regional and inter-sectoral differences and will shed new light on cross-country evidence: the intensity of market competition among firms.

²If anything, Italian labor and tax legislation provides promotive exemptions for the South. As far as the rigour of law enforcement is concerned, there is no reason to believe that South-Italian bureaucrats are uniformly more corrupt than their North-Italian counterparts (see Del Monte/Papagni, 2001)

The reasoning is as follows: A firm which operates in the underground economy can buy its inputs, in particular labor, at a lower price (by avoiding payroll taxes, not complying with safety and health standards, etc.), thereby reducing its variable cost relative to a firm in the official economy. The underground firm can pass on its savings to consumers, which will reduce market prices, and as a result its competitors' profits fall. Thus, the official firm is put at a competitive disadvantage, and may have to choose between operating underground as well, or going out of business. The keener is competition, the higher is the pressure to reduce costs, and the more likely are underground activities to spread in the industry.

This reasoning has some parallels in Shleifer's (2001) argument that competition may promote unethical behavior (e.g. child labor, corruption, etc.). However, the trade-off I consider here is not one between cost savings and the firm owner's private utility of ethical behavior, but the trade-off between cost savings and the threat of detection. While in my model firm owners do not take any moral considerations into account, they face a risk of being detected and fined by the tax authority, a feature that plays no role in Shleifer's (2001) argument.

To my knowledge, there is only one paper that relates the shadow economy to market competition, namely Goldberg/Pavcnik (2003). This paper asks whether we should expect the informal sector in developing countries to expand in response to trade liberalization programs (i.e. to an increase in foreign competition on domestic markets). However, in their model all firms behave as price takers, and so there is no room for strategic interaction among firms which is crucial in my set-up.

In the model proposed in this chapter, the term competition refers to market power, with the source of market power being product differentiation. Product differentiation is a primitive of the model (i.e. I do not consider the possibility of firms choosing their position in product space, or choosing to collude, or any other form of endogenous determination of competition), and I will claim that more intense competition (in the sense that products are closer substitutes, and so market power is lower) implies a larger underground sector.

Anecdotes support the view that the underground economy may expand in response to keener competition. For example, the head of the Austrian Federal Guild of the Construction Industry (Bundesinnungsmeister des Baugewerbes), Mr. Johannes Lahofer, explains the high incidence of shadow economic activity in his industry by recent changes in the way public building contracts are assigned, referring in particular to the introduction of compulsory tenders, and the obligation to assign the contract to the lowest-price bid. (article in "Kurier" of October 4, 2004)

These new regulations prevent local authorities from discriminating against certain firms (and favoring others) for reasons not related to the price offered, and forces them to take all interested construction firms into account when offering a building contract. Applying the concept of competition used in this chapter, we can say that the new laws rendered the construction industry more competitive by imposing full substitutability of all firms from the point of view of the (institutional) buyer. The ensuing increase in shadow-economic activity in the construction industry is therefore consistent with the line of reasoning laid out above, which suggested precisely that outcome.

The chapter proceeds as follows: In Section 3, I present a simple oligopoly model where each firm first decides whether to enter and whether to operate in the official or underground economy; then, competition in the product market takes place. In the product market, each firm chooses the price for its output, and competition will be called "intense" if consumers perceive the firms' products as close substitutes, i.e. if the rivals' price choices have a large impact on each firm's own demand.

My main result will be that firms operating in a more competitive industry will be more likely to operate in the underground economy. The reason is that competition reduces profit margins in the official sector faster than in the underground sector, thus inducing firms to switch to the underground sector. This formalizes the argument loosely stated above, namely that competition will increase the temptation to go underground.

Finally, Section 3 presents cross-country evidence on the impact of entry and competition characteristics on the size of the underground economy in various OECD, transition and developing countries. While the reliability of data on the underground sector is a controversial issue, my regression results indicate that more intense competition is indeed correlated with a higher incidence of shadow-economic activity, thus lending support to my model predictions.

The Model

Description of the Game

I will consider two types of agents: firms and the tax authority. Their behavior and decisions are characterized as follows.

The Firms

Suppose there is a (very large) pool of potential firms in a specific industry. These firms are ex-ante perfectly identical, and play the following two-stage game:

Stage 1: Each firm decides (simultaneously with all other firms) whether to enter the official economy, or to enter the underground economy, or to stay out.³ If the firm stays out, its outside option yields payoff zero.

The choice between the official and the underground economy is irreversible, and I model it as one between two different "production technologies", which are characterized as follows:

(i) Production costs: If firm i operates in the official economy, its total production cost as function of its output q_i is

$$C_o(q_i) = c_o q_i + C_E \quad (3.1)$$

while the total production cost of a firm j operating in the underground economy is

$$C_u(q_j) = c_u q_j \quad (3.2)$$

Denote by $C_E \geq 0$ the entry-regulation cost (red tape) of the official firm, which has to be sunk at stage 1 in order to enter the official economy. Let C_E be smaller than monopoly profits of an official firm, so that the industry is viable. Assume for simplicity that there is no other fixed cost of entry in either sector.

³Note that I treat the decision to operate in the underground economy as an "all-or-nothing" choice, i.e. I do not allow for a single firm to "split" its operations between the official and the underground sector. One may think of these operations as regarding a single (and indivisible) project or activity rather than a "firm" in the broad sense, as the "firms" in this model will only be active for one period (namely stage 2 of the game), after which the game is over and everybody shuts down.

The term $c_o \in (0, 1)$ denotes (constant) marginal production cost of the official firm, while $c_u < c_o$ represents marginal cost when operating in the underground economy. The wedge between c_o and c_u can have different sources: if the firm operates in the underground economy, it can avoid payroll taxes for its workers, can defy environmental or other regulations which increase the cost of production, and avoid the administrative costs associated with tax compliance itself (like keeping records, registering workers with the social security authority etc.).

(ii) Auditing: every firm will be audited by the tax authority with a probability α (where α is common knowledge among all firms). If audited, an agent operating in the underground economy will be detected with certainty and has to pay a fine F ; for an agent who operates in the official economy, the audit will remain without consequences, i.e. I assume that the tax authority never makes mistakes (see section 3 for a discussion of the tax authority and the properties of α and F).

Stage 2: Given that at stage 1, a total number n of firms entered the industry, out of which a share $1 - \mu$ firms decided to operate in the official economy (while μ operate in the underground economy), at stage 2 the firms will simultaneously choose prices. Then, markets clear, and profits are realized; the tax authority audits a fraction α of all firms, and the underground firms that are caught will be convicted to pay the fine $F > 0$.

Competition among the firms is imperfect in the sense that goods are horizontally differentiated, and each firm produces one variety.

Consumers' valuation for a variety does not depend on how this variety was produced,

i.e. whether it was produced in the official or in the underground economy: consumers may not be able to verify how the good was produced, or if they know, they do not perceive any (vertical) quality difference between goods in the official and the underground sector.⁴

Specifically, consumer demand for variety i , q_i , is characterized by

$$q_i(p_i, p_{-i}) = \max \left\{ \frac{1}{n} \left(1 - p_i(1 + \gamma) + \frac{\gamma}{n} \sum_{j=1}^n p_j \right), 0 \right\} \quad (3.3)$$

where p_i is the price chosen by firm i , p_{-i} is the vector of competitors' prices $(p_1, \dots, p_{i-1}, p_{i+1}, \dots)$ and n is the total number of firms operating in the market.

The parameter $\gamma \geq 0$, which will be crucial to the analysis, measures the (symmetric) degree of substitutability (and hence the intensity of competition) between any two varieties i and j ; if $\gamma = 0$, the two varieties are completely independent (hence each firm behaves as a monopolist facing demand $q_i(p_i) = \frac{1}{n}(1 - p_i)$), if γ is large, the two varieties are perceived as close substitutes (and hence competition between the two firms will be very fierce).

This demand function is linearly decreasing in own price, linearly increasing in the average price level (i.e. competitors' prices), and normalized by n , the total number of varieties in the industry. This function has the advantage of being algebraically convenient, and allows us to capture "competition" (in the sense of sensitivity of own demand to rivals' prices) in a single, exogenous, parameter.

Among the special properties of these demand functions (3.3), note that the aggregate

⁴Note that this assumption also implies that consumers do not face any risk of consuming goods produced in the underground economy, i.e. I exclude the possibility of joint legal responsibility of consumer and producer once a firm in the underground economy is caught.

demand

$$Q = \sum_{i=1}^n q_i = v - \frac{1}{n} \sum_{i=1}^n p_i$$

does not depend on the degree of substitution among the products, γ , and that in the case of price symmetry, i.e. $p_i = p$ for all $i = 1, \dots, n$, aggregate demand does not change with the number of products n existing in the industry.⁵

The Tax Authority

I make the following assumptions:

The tax authority can only intervene at the end of stage 2 of the game (i.e. after firms produced and sold their output), but not at stage 1.⁶ At stage 2 of the game, the tax authority cannot directly observe the prices charged (and the quantities sold) by the firms on the final good market.⁷

The tax authority can enforce full payment of the fine, i.e. no partial or total default is possible. This implies that: (i) firms must have sufficient assets to cover the fine⁸, and (ii) the tax authority can seize all assets of the underground firms it detects⁹.

Both the audit probability α and the fine F are exogenous from the point of view of a

⁵For a discussion of the derivation and properties of this demand function, see Shubik/Levitan (1980) and Motta (2003).

⁶Recall that firms entering the official sector pay entry regulation cost C_E . Thus, their number and identity becomes immediately observable to the tax authority. Underground firms, however, cannot be distinguished from non-entering firms, until they become active, i.e. produce (and sell) a strictly positive quantity at stage 2.

⁷As we will see later, the prices charged by underground firms will differ systematically from those of official firms; thus, if the tax authority could observe these prices, it could easily identify the underground firms, and the detection probability would have to be 1.

⁸This will be the case if firms have revenues from activities outside of the industry considered in this model, or if the fine is (partially) non-pecuniary (e.g. prison sentences, reputational penalties).

⁹This assumption may not always be satisfied in practice, where underground firms may just shut down their premises and "disappear" when they are caught.

single firm. This assumption implies that:

(i) The audit probability α does not vary with a firm's output; in particular, an underground firm is not more likely to attract the tax authority's attention because it produces more.

(ii) F is independent of the incriminated firm's output and profits, i.e. the fine is the same for all firms, no matter what the amount of taxes evaded is.

However, I allow the expected fine, αF , to vary with the aggregate share of underground firms in the industry, μ . In particular, assume that $\alpha F(\mu)$ is some continuous function of μ . I do not make any assumptions about the exact shape of $\alpha F(\mu)$ (it may be constant, increasing, decreasing, or non-monotonic in μ) and about the determinants of the tax authority's behavior (such as resource or informational constraints, revenue targets, etc.) that could give rise to such a function.¹⁰

Solution of the Game

I will now identify the subgame-perfect pure-strategy equilibria of the game described above.

Equilibrium in the Product Market (stage 2)

Moving backwards, let us first solve for the equilibrium of the price-choice stage. Given that at stage 1, a total number n of firms entered the market, out of which $n(1 - \mu)$ firms

¹⁰We will see below that the properties of $\alpha F(\mu)$ are decisive for the type of equilibria that can arise in this game.

decided to operate in the official economy (while $n\mu$ operate in the underground economy), a firm i which decided to operate in the official economy will maximize its gross profits as follows:

$$\max_{p_i} \{ (p_i - c_o) q_i(p_i, p_{-i}) \} \quad (3.4)$$

while a firm j that opted for the underground economy has to solve

$$\max_{p_j} \{ (p_j - c_u) q_j(p_j, p_{-j}) \} \quad (3.5)$$

where $q_i(p_i, p_{-i})$ and $q_j(p_j, p_{-j})$ are defined as in equation (3.3)¹¹.

Let us impose symmetry among the $n(1 - \mu)$ firms which operate in the official economy (i.e. all firms in this sector charge the same price, p_o) and among the $n\mu$ firms that operate in the underground economy (which will all charge p_u). Thus,

$$\sum_{l=1}^n p_l = n(1 - \mu)p_o + n\mu p_u.$$

I solve for p_o^* and p_u^* (the equilibrium prices charged by the typical firm in the official and underground economy, respectively) to obtain:

$$p_o^*(n, \mu) = \frac{(2 + 2\gamma - \frac{2}{n})(1 + c_o(1 + \gamma - \frac{2}{n})) + (c_u - c_o)\gamma(1 + \gamma - \frac{2}{n})\mu}{(2 + 2\gamma - \frac{2}{n})(2 + \gamma - \frac{2}{n})} \quad (3.6)$$

and

$$p_u^*(n, \mu) = \frac{(2 + 2\gamma - \frac{2}{n})(1 + c_u(1 + \gamma - \frac{2}{n})) + (c_o - c_u)\gamma(1 + \gamma - \frac{2}{n})(1 - \mu)}{(2 + 2\gamma - \frac{2}{n})(2 + \gamma - \frac{2}{n})} \quad (3.7)$$

¹¹Note that at this stage, i.e. conditional on having opted for the underground sector, the threat of detection has no influence on the firm's behavior anymore. This is due to the assumption that αF is independent of $q_j(p_j, p_{-j})$, implying that second-stage (price) choices will be unaffected by the expected fine.

Now, the first-order conditions imply that the equilibrium quantities sold by each firm are

$$q_o^*(n, \mu) = \max \left\{ \frac{1}{n} (p_o^* - c_o) \left(1 + \gamma - \frac{\gamma}{n} \right), 0 \right\} \quad (3.8)$$

and

$$q_u^*(n, \mu) = \max \left\{ \frac{1}{n} (p_u^* - c_u) \left(1 + \gamma - \frac{\gamma}{n} \right), 0 \right\} \quad (3.9)$$

so that gross profits in the price-choice equilibrium are

$$\Pi_o(n, \mu) = \begin{cases} \frac{1}{n} (p_o^* - c_o)^2 \left(1 + \gamma - \frac{\gamma}{n} \right) & \text{if } q_o^* > 0 \\ 0 & \text{if } q_o^* = 0 \end{cases} \quad (3.10)$$

and

$$\Pi_u(n, \mu) = \begin{cases} \frac{1}{n} (p_u^* - c_u)^2 \left(1 + \gamma - \frac{\gamma}{n} \right) & \text{if } q_u^* > 0 \\ 0 & \text{if } q_u^* = 0 \end{cases} \quad (3.11)$$

The following Lemma highlights some of the properties of the product market equilibrium:

Lemma 12 *In equilibrium, firms operating in the underground economy:*

- (i) *charge a lower price than firms in the official economy, i.e. $p_u^* < p_o^*$*
- (ii) *have higher mark-ups than firms in the official economy, i.e. $p_u^* - c_u > p_o^* - c_o$*
- (iii) *make larger gross profits than official firms, i.e. $\Pi_u > \Pi_o$*

Proof: see Appendix A

Note that these price and profit effects described in Lemma 12 are entirely driven by the fact that underground firms produce at a lower marginal cost than official firms. The resulting cost advantage is partly passed on to consumers (through lower prices), partly retained by the underground firms (through higher markups).

Equilibrium at the Entry-Stage of the Game (stage 1)

At stage 2 of the game, firms take the total number of firms in the industry, n , and the number of firms in the underground economy, $n\mu$, as given. Hence, equilibrium prices, $p_o^*(n, \mu)$ and $p_u^*(n, \mu)$, as well as gross profits, $\Pi_o(n, \mu)$ and $\Pi_u(n, \mu)$, are all functions of n and μ . Both n and m will be determined simultaneously at stage 1 of the game.

When deciding whether to enter the official or the underground economy (or to stay out), firms will correctly anticipate the equilibrium of the second stage of the game. They will also take into account the fixed entry-regulation cost, C_E , that has to be paid when entering the official economy, and hold it against the risk of being detected and fined when entering the underground economy¹².

Any equilibrium of the first stage will have to satisfy the following conditions:

- (free entry) None of the inactive firms could make strictly positive net profits by entering the industry;
- (breaking even) None of the active firms makes losses (i.e. none of them would strictly prefer to remain inactive);

¹²Note that both C_E and $\alpha F(\mu)$ are sunk at stage 2, and therefore do not matter for the price choice.

- (free sector choice) None of the firms active in one sector could make higher net profits by switching to the other sector.

More formally, we can define the subgame-perfect, pure-strategy equilibria of the first stage of the game as a pair (μ^*, n^*) such that¹³:

(i) "Coexistence Equilibria": If firms will be active in both the official and the underground sector of the industry, i.e. if $\mu^* \in (0, 1)$, then (μ^*, n^*) must solve

$$\Pi_o(\mu, n; \cdot) - C_E = \Pi_u(\mu, n; \cdot) - \alpha F(\mu) = 0$$

(ii) "Pure Official Equilibria": If all active firms operate in the official sector, and no firm is active in the underground economy, i.e. $\mu^* = 0$, then $(\mu^* = 0, n^*)$ must solve

$$\Pi_u(\mu, n; \cdot) - \alpha F(\mu) \leq \Pi_o(\mu, n; \cdot) - C_E = 0$$

(iii) "Pure Underground Equilibria": If all active firms operate in the underground sector, and no firm is active in the official economy, i.e. $\mu^* = 1$, then $(\mu^* = 1, n^*)$ must solve

$$\Pi_o(\mu, n; \cdot) - C_E \leq \Pi_u(\mu, n; \cdot) - \alpha F(\mu) = 0$$

Which of these equilibria will actually arise depends on how the threat of detection plays out against the higher marginal cost and entry cost of operating in the official economy.

Proposition 13 *The game described above*

(i) has at least one subgame-perfect pure-strategy equilibrium (this may be a coexistence equilibrium, or a pure official or pure underground equilibrium);

¹³To simplify the analysis, I will treat both n and $n\mu$ as real numbers, even though they are of course constrained to be positive integers. Thus, the equilibria described and analyzed in the following are in fact just quasi-equilibria.

(ii) may have multiple equilibria (both pure equilibria, or multiple coexistence equilibria, or any combination of pure and coexistence equilibria).

Proof: see Appendix A

Intuitively, if the expected fine is very high, firms may be fully deterred from entering the underground economy, and we will only see official firms operating. Conversely, if enforcement is close to inexistent, then all firms will operate underground. If, instead, the expected fine is somewhere in-between, so that not all firms will want to be in the same sector, we obtain coexistence equilibria.

Comparative Statics

Recall that our objective was to evaluate the impact of intensity of competition, represented by parameter γ , on the size of the underground economy. For this purpose, let us restrict attention to the coexistence equilibria, that is equilibria where firms are active in both the official and the underground sector of the industry.

Thus, let $\alpha F(\mu)$ be such that there exists at least one pair (μ^*, n^*) , where $\mu^* \in (0, 1)$ and $n^* > 1$, solving the equilibrium conditions

$$\Pi_o(\mu, n; \gamma, \cdot) - C_E = \Pi_u(\mu, n; \gamma, \cdot) - \alpha F(\mu) = 0$$

Starting from such an equilibrium, suppose we let the competition parameter γ vary slightly. Then, this change will affect the firms' gross profits in both sectors (gross profits will decrease if γ increases), and so our equilibrium conditions will no longer be satisfied at the initial solution (μ^*, n^*) . Thus, μ and n will have to adapt accordingly to allow the

industry to settle at a new equilibrium. Proposition 14 tells us in which direction this change in μ will go.

Proposition 14 *If the coexistence equilibrium is stable, then the equilibrium share of firms operating in the underground economy, μ^* , is increasing with respect to the intensity of competition γ . In other words, as the industry becomes more competitive, firms will be more likely to operate in the underground economy.*

Proof: see Appendix A

Discussion

The result derived above relies on two key features of the model:

(i) Operating in the underground economy allows firms to produce at lower marginal cost than firms in the official sector;

(ii) Firms' product-market (i.e. price) choices can be separated from the entry and sector choices (sequential decision making) and from all considerations regarding the risk of detection (the expected fine is independent of an underground firm's price or profits)

When deciding which sector to enter, firms face a trade-off: in the official sector, they make lower gross profits than underground firms, and have to pay the entry-regulation cost; in the underground economy, however, they face the risk of detection and punishment.

If, in equilibrium, firms are active in both the official and the underground economy, then the share of firms in either sector will exactly balance this trade-off. Now, as competition becomes more intense, markups in both sectors of the industry will drop, but markups in the

official sector will drop faster, thus shifting the balance in favor of the underground economy.

This result is robust to several modifications of the setup:

- allowing for product market competition in quantities instead of prices
- introducing (flat) taxation of official firms' profits; this creates additional incentives for firms to go underground
- introducing additional fixed cost (physical setup costs) in both sectors on top of the entry-regulation cost that firms have to pay to enter the official economy
- allowing the detection probability α to depend on these physical setup costs (to incorporate the idea that the larger the facilities required for production, the more "visible" a firm will be, and the more difficult it will find it to hide its operations from the tax authority)
- parameterizing market size (where market size is captured by the intercept of the demand function, and was set to 1 in the analysis above)
- allowing for different functional forms of the demand function (note that both the existence of equilibria and the comparative statics rely on the continuity of gross profits in all parameters, and the signs of the corresponding partial derivatives, not on the specific functional form assumed for demand)

Some of the assumptions in the model may seem strong and deserve a more thorough discussion:

(i) Recall that in this model, the term competition refers to a firms' ability to price above marginal cost. This is not the only sense in which this term can be used; "competition" may refer to both market structure and market outcome. As for market structure, we may think of

competition as being restricted by the presence of entry barriers (in particular administrative barriers like trade licenses), which reduce the number of firms that can enter the (official) industry, and which may entice entrepreneurs into "bypassing" them by offering their goods or services without the required permits. Thus, if we equate low competition with high entry barriers, we should expect low competition to be associated with a high incidence of shadow-economic activity.

However, it is important to distinguish between the entry aspect and firms' behavior in the market after entry. Once entry decisions have been made, firms may compete fiercely in the sense that they charge prices close to marginal cost, or they may enjoy market power, that is they may be able to raise prices well above marginal cost without losing all their buyers to their competitors.

(ii) The assumption that the fine F is independent of the incriminated firm's output and profits may seem unrealistic, because, in practice, enforcement authorities tend to tailor the punishment "to fit the crime". For instance, tax authorities may set fines according to an estimate of the amount of taxes evaded.

Yet, the scope for variable fines may be limited for several reasons. First, to make an estimate of taxes evaded, the firms's profits would have to be verifiable, which may not always be the case.¹⁴ Second, apart from evading taxes, avoiding compliance with labor and environmental laws may be an important motivation for operating underground. Yet, this damage is more difficult to quantify and to translate into monetary terms, and so fixed-fee

¹⁴Recall the assumption that prices and quantities are not directly observable by the tax authority: it could be the case that even after investigating an underground firm and finding the evidence necessary for conviction, the firm's output and profits remain non-verifiable (though potentially observable).

punishments are more likely applied to these types of infringements.

(iii) Another feature of the model that may raise concerns is the "all-or-nothing" nature of the sector choice. In practice, there are many firms that split their operations between the two sectors, a decision that cannot arise in the model considered so far.

However, a simple illustration will show that my model prediction holds good even in a very different setting, where I allow for both types of operations within the same firm: Consider a perfectly competitive industry, where each firm behaves as a price taker. Each firm chooses the total output q it wants to sell at the going market price. Each unit of output can be produced in one of two ways: either "officially", i.e. using declared inputs, in which case marginal cost is some convex function $c_o(q)$; or "underground", i.e. using undeclared inputs, which is associated with convex marginal cost $c_u(q)$.

Interpret $c'_o(q) > 0$ as an inherent property of the production technology, which may be due to short-run capacity constraints, while $c'_u(q) > c'_o(q)$ reflects the combined effect of the technological constraints and the threat of detection and punishment, which I assume to be increasing in the underground output. Let $c_u(q=0) < c_o(q=0)$, so that, for very low levels of output, producing "underground" is unambiguously more profitable for the firm.

Suppose that the two marginal cost curves intersect at some output level, call it $q^* > 0$, so that for all $q > q^*$, the benefits of underground production (payroll tax evasion etc.) are outweighed by the increasing risk of detection. Then, the firm's short-run supply curve is the lower envelope of these two marginal-cost functions, that is, the firm will produce part of its output (up to q^*) using undeclared inputs, and any q exceeding q^* using declared inputs.

In equilibrium, our firm will produce the q that solves

$$p = C'(q) \text{ where } C'(q) \equiv \begin{cases} c_u(q) & \text{for } q \leq q^* \\ c_o(q) & \text{for } q > q^* \end{cases}$$

Suppose that initially the equilibrium price was high enough to induce the firm to produce more than q^* , i.e. to have some positive official output. Next, assume that there is a negative shock to the equilibrium price, i.e. the price falls to $p' < p$. This could be the result of a drastic cut in tariffs which allowed more efficient foreign firms access to the domestic market, or some other exogenous event that makes the environment for domestic firms more "competitive".

Then, the firm will reduce its output to the level which solves $p' = C'(q)$. Note that the first units of output that will be "crowded out" are the officially produced ones; only when output falls even below q^* will the firm start reducing its underground operations as well. In either case, the ratio of underground to official output will increase, and if all firms are symmetric, then the industry-wide underground economy will have grown in size.

Empirical Evidence

Description of the Data

In the following, I will perform a cross-country analysis using a panel that covers 48 countries (OECD, transition and developing countries) for the years 1995 to 2000.¹⁵ The

¹⁵The countries are: Argentina, Australia, Austria, Belgium, Bolivia, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Ecuador, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea (Rep.), Malaysia, Mexico, Netherlands, New Zealand, Norway, Peru, Philippines, Poland, Portugal, Russian Federation, Singapore, Slovakia, Spain, Sweden, Switzerland, Thailand, Ukraine, UK, USA, and Venezuela.

underlying data were drawn from several sources: the "Global Competitiveness Report" (GCR), which is published annually by the World Economic Forum; Schneider's (2002a, 2002b, 2003) estimates of the underground economy as percentage of official GDP; and the World Bank's "World Development Indicators", "Doing Business" and "Labor Regulation" databases.

The measures of most variables contained in the Global Competitiveness Report are based on the results of the Executive Opinion Survey, which asks some 4,000 top and middle managers in the surveyed countries for a personal assessment of the variables of interest.¹⁶ Each respondent assigned an integer from 1 to 7 to each of the questions contained in the survey, and the Global Competitiveness Report reports the *average* response for each variable and country; unfortunately, I do not have access to the distribution of responses underlying the computation of the average.

The two measures for the size of the underground economy used in the following deserve some more discussion. In terms of the model of Section 3, these measures are meant to proxy μ^* , the equilibrium share of firms in the underground economy¹⁷. Of course, the very nature of the subject matter makes it difficult to quantify it. There are two major approaches to measuring the underground economy: the "indirect" approach, which relies on national account statistics to infer the approximate size of the shadow economy, and the "direct"

¹⁶ Note that the managers' assessments may not be representative for the opinion held by the general public in their country; however, for the purposes of our analysis, what I am interested in is precisely the perception of firm-level decision-makers, as they are the ones who choose whether or not their firm will operate in the underground economy.

¹⁷ Note, however, that the indicator "size of the underground economy as percentage of official GDP" does not exactly correspond to μ^* , which represents the relative number of firms in the underground sector, not their relative output.

approach, which is based on information obtained through interviews.

(a) Currency-Demand Approach

The data series in Schneider (2002a, 2002b, 2003) are of the "indirect" type, i.e. they are estimates of the size of the underground economy as share of official GDP using the currency-demand approach. Unfortunately, they are only available for 3 out of the six periods covered by the GCR (namely 1995, 1998 and 2000); moreover, for the year 1998, this data set only covers the OECD countries (i.e. 21 out of the full set of 48 countries).

(b) Electricity Approach

To complement the analysis, I will therefore repeat all estimations using a different measure of the underground economy, based on the so-called "physical input" method (or Kaufmann-Kaliberda method). To measure overall (official and unofficial) economic activity, Kaufmann and Kaliberda (1996) assume that electricity consumption is the single best physical indicator of overall economic activity. With the electricity-GDP elasticity usually being close to one, the difference between the growth of official GDP and the growth of electricity consumption can be attributed to the growth of the shadow economy. (Source of GDP and electricity consumption data: World Bank World Development Indicators)

While this measure of the underground economy is certainly less sophisticated than the currency-demand measure, it has two main advantages:

- (i) data are available for all six periods covered by the GCR, and
- (ii) the measure is directly derived from consumption of a physical input, and so we do not have to worry about interactions between the variables used to construct the currency-

demand estimates¹⁸ and the explanatory variables used in the regressions, which are explained below.

Table 1 summarizes the variables used in the following analysis, the data sources, and the model parameters that these measures should capture. For a more detailed description of the individual measures, see Appendix B.

¹⁸These variables are: ratio of cash holdings to current and deposit accounts, weighted average tax rate, proportion of wages and salaries in national income, interest paid on savings deposits, and per-capita income.

Table 1: *List of Variable Names, Sources, and Model**Parameters That These Variables Correspond to*

Variable Name	Source	Model
Underground Economy:		
1. currency-demand approach	Schneider (2002, 2003)	μ^*
2. electricity approach	own calculations	μ^*
Low corruption	GCR	$\alpha F(\mu)$
Labor regulation	WB "Labor Regulation"	$c_o - c_u$
Payroll tax rate	GCR	$c_o - c_u$
Income tax rate	WB WDI	control
Corporate tax rate	WB WDI	control
VAT	WB WDI	control
Cost of Start-up	WB "Doing Business"	C_E
Local Competition	GCR	γ

GCR = Global Competitiveness Report, WB = World Bank,

WDI = World Development Indicators

Table 1A in Appendix B presents summary statistics of the raw data described above, while Table 1B in Appendix B reports the correlation coefficients of all variables.

Now, recall that my model explained the equilibrium share of firms in the underground economy, μ^* , as a function of marginal costs c_o and c_u , red-tape cost C_E , rigor of enforcement,

$\alpha F(\mu)$, and intensity of competition γ . In particular, Proposition 14 predicts that the size of the underground economy should be increasing in the intensity of γ .

In what follows, I will examine if this relationship identified in my model is confirmed or rejected by cross-country analysis. A caveat is in order: the theoretical unit of interest is the individual firm operating in a specific industry, while the available data report the variables of interest (in particular the share of the underground economy and the intensity of competition) at the country-level, aggregating over all industries. Thus, my empirical analysis should be viewed as exploratory analysis motivated by theory, rather than an actual test of my model.

Regression Results

I will estimate the following regression:

$$UNDECY = \beta_0 + \beta X + \beta_1 COMP + \varepsilon$$

where the dependent variable, *UNDECY*, is "underground economy", β_0 is a constant, X is the vector of controls ("low corruption", "labor regulation", "payroll tax %", "income/corporate tax %", "VAT", "cost of start-up"), *COMP* is "competition", and ε is the error term.

Note that two of these variables, namely "labor regulation" and "cost of start-up", are only available for one year. Yet, as these two series measure structural characteristics which are not likely to change a lot over a time span of 6 years, they were included as constants in the estimations. This means that the standard panel-analysis techniques, like fixed and

random effects analysis, cannot be applied appropriately to the data set, because the country-fixed effects will already be picked up by these two other variables; thus, I will restrict myself to performing pooled least squares regressions on the stacked data set.

Now, recall from Section 3 that there are good reasons to believe that a country's tax rates are to some extent endogenous, i.e. correlated with the error terms of its size of the underground economy: If tax rates are increased in response to an expansion of shadow economic activity, then the causality between tax rates and the underground economy would be reversed, and our coefficient estimates would be inconsistent. (see Johnson et al (1998)).

In order to avoid inconsistent estimation of the parameters of interest, I will therefore use a two-stage instrumental variable procedure. At the first stage, I predict each country's tax rates by their instruments, and then use these predicted values as regressors in our final regression. To do this, I assign to each country i one neighboring country j , whose tax rates will be used as instruments for i 's tax rates.

Table 2 in Appendix B reports the results of this first-stage regression of tax rates on their instruments (including a country-specific constant).

With the R^2 of these regressions ranging from 0.53 to 0.95, they yield reasonably good predictions for the endogenous tax rates. Moreover, it seems unlikely that one country's tax rate policies could be influenced by changes in the size of a neighboring country's underground economy. Thus, the neighboring country's tax rates qualify as instruments for the endogenous regressors.

Now, we can turn to the second stage of the regression analysis: the following table shows

the results of two different OLS regressions on the stacked data, where I first perform an OLS regression on the estimates for the underground economy derived from the currency-demand approach, and then an analogous OLS regression with the dependent variable resulting from the electricity approach calculations. I obtain the following results, where the terms in brackets are significance levels:

Table 3: *Second-Stage OLS Regression Results*

Sample: 1995 -2000 (6 years)		
OLS I: Dep. Var.: "Underground Economy" (currency-demand approach)		
OLS II: Dep. Var.: "Underground Economy" (electricity approach)		
	OLS I	OLS II
constant	22.95 (0.0039)	38.95 (0.0000)
low corruption	-4.35 (0.0003)	-5.32 (0.0000)
labor regulation	12.01 (0.0000)	12.02 (0.0000)
payroll tax % IV	-0.26 (0.0017)	-0.34 (0.0000)
income tax % IV	-0.23 (0.0535)	0.002 (0.9738)
corporate tax % IV	-0.24 (0.1776)	-0.34 (0.0021)
VAT % IV	0.49 (0.0067)	0.25 (0.0400)
Cost of Start-up (% GNI)	-0.06 (0.0298)	-0.08 (0.0000)
Competition	3.96 (0.0011)	1.80 (0.0118)
country-years	102	248
Adjusted R^2	0.48	0.53

First of all, we see that the most important explanatory variables are labor regulation, corruption, and the intensity of competition. The estimated coefficients of these three variables display the expected signs and are highly significant under both specifications. This indicates that good governance and rigorous enforcement play an important role in keeping the underground sector small. Likewise, a highly regulated labor market seems to encourage the expansion of the underground economy.

We also see that a one-point increase in the competition index will lead to an increase in the size of the underground economy of 1.8 to 4 percentage points. Recall that this is consistent with the predictions of my model from Section 3, where we concluded that more intense competition leads to a larger underground sector.

Finally, it is striking that the variables that one would expect to explain most of the variation in the dependent variable actually have little explanatory power: the coefficients on "cost of starting a new business" and the various tax rates are small in size, partly insignificant, and most of the time even display counterintuitive signs.¹⁹

Robustness

As a first step, let us check if the results are robust to the introduction of regional dummies (4 dummies, one for OECD countries, Central and Eastern Europe, Latin America, and Asia). As reported in Appendix C, Table 4, the coefficients on the variables of interest do not change sign and remain significant if they were so in the base specification. Also

¹⁹The insignificance of the tax-rate variables may seem particularly surprising. The result is also robust to including *squared* tax rates (which would account for concave effects of the tax rates). Yet, the insignificance of the tax rates is in line with the findings of Friedman et al (1999) and Johnson et al (1998).

note that the coefficients on the regional dummies are almost always significant as well, and that Latin America and Asia tend to have somewhat higher point estimates than OECD and CEE countries.

This last result could point to heterogeneity among the countries in the sample with respect to the way that the explanatory factors influence their underground economy. To address this question, the sample was split into four sub-samples (OECD countries, Central and Eastern Europe, Latin America, and Asia) and a separate regression was run for each of the countries in this group.

The results (not reported here) are broadly consistent with the ones obtained from the base specification, but most coefficients are no longer significant, in particular for the Latin American and Asian sub-sample, while they remain highly significant for the OECD sample (with smaller point estimates, though). This is probably due to the fact that there are very few data points for CEE, Latin America and Asia once I treat them as separate samples, and so the estimates become imprecise.

As a next step, the regressions reported in Table 3 were re-estimated dropping one variable at a time so as to check that the specification is robust (results not reported here). As it turns out, coefficient estimates never change sign and always remain significant if they were so in the base specification, with one exception: if "low corruption" is dropped, then the coefficient on "competition" will become insignificant, and even negative when underground economy is being measured by the electricity-variable.

Therefore, it might be insightful to re-run the regressions exchanging one variable at a

time by a different measure for that very same variable, as far as such alternative measures are available²⁰. Now, when substituting "low corruption" by the Corruptions Perception Index (published annually by "Transparency International") or the governance measures of the World Bank's "Governance Indicator" database ("perceptions of corruption" and "rule of law"), we see again that the coefficient on "competition" turns negative and insignificant (results not reported here).

Yet, when substituting "competition" by "average tariff rate" (as a measure of the presence of foreign competition curbing market power of domestic firms), then the sensitivity to the measure of corruption vanishes. Thus, the problem may be due to the fact that the measure for "competition" is flawed: there could be an endogeneity problem due to survey respondents who indicated that "competition" is high because the underground sector in this industry is high, and competition from this sector is perceived as particularly tough.

Thus, the base regression was repeated, but this time instrumenting for "competition" by "fuel, ores, and metal exports" (as a measure of natural domestic rents, which should be negatively correlated with "competition"), "average tariff rate" (as before), "effectiveness of antitrust policies" (which should immediately affect local competition), and the log of population (as a measure of domestic market size).

As Table 5a shows, all regressors of the first-stage regression are significant and have the expected sign. Then, Table 5b shows the second-stage regression, where "competition" was instrumented for. We see that the coefficient on "competition" is still positive and

²⁰For instance, replacing "labor regulation" by the corresponding measures provided by the World Bank's "Doing Business" database (Conditions of Employment Index, Flexibility of Hiring/Firing Index) does not change the results, nor does the replacement of "cost of starting a new business" by the corresponding variable contained in the GCR (results not reported here).

significant, though somewhat smaller than in the base regression of Table 3.

To summarize, once I instrument for the key regressor, the results presented in Table 3 seem to be robust to various changes in specification, like introducing regional dummies, splitting the sample into sub-samples, and dropping or replacing explanatory variables.

Conclusion

The purpose of this chapter is to study a novel rational for the variations in the share of the underground economy which sheds new light on cross-country evidence: the intensity of market competition among firms.

To address this issue, I first developed a simple oligopoly model of price competition with differentiated goods to analyze equilibrium outcomes of this decision. In this model, the individual firm can freely choose whether to enter the official or the underground sector, and the intensity of competition in the industry is captured by a single parameter that represents the homogeneity of product varieties.

I argued that when deciding which sector to enter, firms face a trade-off: in the official sector, they make lower gross profits than underground firms, and have to pay the entry-regulation cost; in the underground economy, however, they face the risk of detection and punishment. If, in equilibrium, firms are active in both the official and the underground economy, then the share of firms in either sector will exactly balance this trade-off.

My main result is that as the industry becomes more competitive (in the sense that the firms' product varieties become closer substitutes), the firms will be more likely to

operate in the underground economy. We saw that competition reduces profit margins in the official sector faster than in the underground sector, thus increasing the temptation to go underground.

In the following, I performed a cross-country analysis using a panel that covers 48 countries (OECD, transition and developing countries) from 1995 to 2000. Pooled least squares regressions show that ease-of-entry and tax rates are not the only forces behind the emergence of the underground economy (in fact, their contribution is much weaker than one would expect).

Instead, good governance as well as labor market regulation seem to have the largest explanatory power. Moreover, the regression results indicate that more intense competition is indeed correlated with a higher incidence of shadow-economic activity, i.e. the empirical findings corroborate my model predictions.

Several issues are raised by this chapter that deserve further investigation: One key element of the model is the tax authority's behavior, which is taken as given in my model without looking into its determinants. Another issue to investigate are the welfare effects of underground activity. On the one hand, underground firms evade taxes and fail to comply with labor and environmental regulations, thus generating considerable social costs. On the other hand, their presence exerts downward pressure on the prices charged by official firms, which benefits consumers. I would need to make precise assumptions on the weights of these effects in the social welfare function to draw firm conclusions.

Bibliography

- [1] ALM, JAMES/MCCLELLAND, GARY H./SCHULZE, WILLIAM D. (1992): "Why Do People Pay Taxes?", *Journal of Public Economics* 48(1), 21-38.
- [2] ANDERBERG, DAN (2003): "Insiders, Outsiders, and the Underground Economy", *CESifo Working Paper* No. 1048, October 2003.
- [3] ANDREONI, JAMES/ERARD, BRIAN/FEINSTEIN, JONATHAN (1998): "Tax Compliance", *Journal of Economic Literature* 36(2), 818-860.
- [4] ARNDT, SVEN W. (1999): "Globalization and Economic Development", *Journal of International Trade and Economic Development* 8(3), 309-318.
- [5] CEBULA, RICHARD J./SALTZ, IRA S. (2001): "An Empirical Note on Tax Auditing and the Size of the Underground Economy in the United States, 1962-1980", *International Review of Economics and Business* 48(1), 119-124.
- [6] CONESA, JUAN CARLOS/DIAZ MORENO, CARLOS/GALDON SANCHEZ, JOSE ENRIQUE (2002): "Explaining Cross-Country Differences in Participation Rates and Aggregate Fluctuations", *Journal of Economic Dynamics and Control* 26(2), 333-345.
- [7] DE RITA, GIUSEPPE/CAMUSI, MARIA PIA (2003): "The Dynamics of the Underground Economy: An Agenda for Action", *Review of Economic Conditions in Italy* 0(1), 53-73.

- [8] DEL MONTE, ALFREDO/PAPAGNI, ERASMO (2001): "Public expenditure, corruption, and economic growth: the case of Italy", *European Journal of Political Economy* 17(1), 1-16.
- [9] EILAT, YAIR/ZINNES, CLIFFORD (2002): "The Shadow Economy in Transition Countries: Friend or Foe? A Policy Perspective", *World Development* 30(7), 1233-54.
- [10] FRIEDMAN, ERIC/JOHNSON, SIMON/KAUFMANN, DANIEL/ZOIDO-LOBATON, PABLO (1999): "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries", *World Bank Discussion Paper*
- [11] GERVAIS, JEAN-PHILIPPE/LARUE, BRUNO (2001): "Do Reductions in Black Market Exchange Rate Premia Cause Inflation?", *Empirical Economics* 26(3), 525-551.
- [12] GILES, DAVID E. A./CARAGATA, PATRICK J. (2001): "The Learning Path of the Hidden Economy: The Tax Burden and Tax Evasion in New Zealand", *Applied Economics* 33(14), 1857-1867.
- [13] GOLDBERG, PINELOPI K./PAVCNIK, NINA (2003): "The Response of the Informal Sector to Trade Liberalization", *NBER Working Paper* 9443, January 2003.
- [14] JOHNSON, SIMON/KAUFMANN, DANIEL/ZOIDO-LOBATON, PABLO (1998): "Regulatory Discretion and the Unofficial Economy", *American Economic Review* 88(2), 387-392.
- [15] KAUFMANN, DANIEL/KALIBERDA, ALEKSANDER (1996): "Integrating the Unofficial Economy into the Dynamics of Post Socialist Economies: A Framework of Analyses and Evidence", *World Bank Policy Research Working Paper* 1691.
- [16] KESSELMANN, JONATHAN R. (1989): "Income Tax Evasion - An Intersectoral Analysis", *Journal of Public Economics* 38 (2), 137-182.
- [17] LACKÓ, MÁRIA (2000): "Hidden economy - an unknown quantity?", *Economics of Transition* 8(1), 117-149.

- [18] LEMIEUX, THOMAS/FORTIN, BERNARD/FRECHETTE, PIERRE (1994): "The Effect of Taxes on Labor Supply in the Underground Economy", *American Economic Review* 84(1), 231-254.
- [19] LEVENSON, ALEC R./MALONEY, WILLIAM F. (1998): "The Informal Sector, Firm Dynamics, and Institutional Participation", *World Bank Policy Research Working Paper* No. 1988, September 1998.
- [20] MOGENSEN, GUNNAR V./KVIST, HANS K./KÖRMENDI, ESZTER/PEDERSEN, SOREN (1995): "The Shadow Economy in Denmark 1994: Measurement and Results", Study No. 3, Rockwool Foundation Research Unit, Copenhagen.
- [21] MOTTA, MASSIMO (2003): "Competition Policy: Theory and Practice", Cambridge University Press, 2003.
- [22] MUMMERT, ANNETTE/SCHNEIDER, FRIEDRICH (2001): "The German Shadow Economy: Parted in a United Germany?", *FinanzArchiv* 58(3), 286-316.
- [23] OECD (2002): "Measuring the Non-Observed Economy - A Handbook", OECD Publications, Paris.
- [24] PESTIEAU, PIERRE/POSSEN, URI M./SLUTSKY, STEVEN M. (1994): "The Penalty for Tax Evasion When Taxes are Set Optimally", Université Catholique de Louvain *CORE Discussion Paper*: 9416, March 1994.
- [25] SCARLATO, MARGHERITA (2001): "Settore sommerso ed equilibri multipli: Uno schema analitico per il Mezzogiorno", *Rivista di Politica Economica*, 91(10), 37-59.
- [26] SCHNEIDER, FRIEDRICH/ENSTE, DOMINIK H. (2000): "Shadow Economies: Size, Causes, and Consequences", *Journal of Economic Literature* 38(1), 77-114.
- [27] SCHNEIDER, FRIEDRICH (2002a): "The Size and Development of the Shadow Economies and Shadow Economy Labor Force of 16 Central and South Ameri-

can and 21 OECD Countries: First Results for the 90s", *mimeo* (available at <http://www.econ.jku.at/Schneider/publik.html>).

- [28] SCHNEIDER, FRIEDRICH (2002b): "The Size and Development of the Shadow Economies and Shadow Economy Labor Force of 18 Asian and 21 OECD Countries: First Results for the 90s", *mimeo* (available at <http://www.econ.jku.at/Schneider/publik.html>).
- [29] SCHNEIDER, FRIEDRICH (2003): "The Size and Development of the Shadow Economies and Shadow Economy Labor Force of 22 Transition and 21 OECD Countries: What Do We Really Know?", *mimeo* (available at <http://www.econ.jku.at/Schneider/publik.html>).
- [30] SHLEIFER, ANDREI (2004): "Does Competition Destroy Ethical Behavior?", *American Economic Review* 94 (2), 414-418.
- [31] SHUBIK, MARTIN/LEVITAN, RICHARD E. (1980): "Market Structure and Behavior", Cambridge, MA, Harvard University Press.
- [32] WORLD ECONOMIC FORUM: "Global Competitiveness Report", Geneva, Switzerland, World Economic Forum, 1996, 1997, 1998, 1999, 2000, 2002.

Appendix A: Proofs

Proof of Lemma 12:

(i) Subtracting $p_u^*(n, \mu)$ from $p_o^*(n, \mu)$ we obtain: $p_o^* - p_u^* = \frac{(c_o - c_u)(\gamma n - \gamma + n)}{(2n + 2\gamma n - \gamma)} > 0$.

(ii) Starting from $0 < n(1 + \gamma)$ which holds by assumption, expand both sides of the inequality to have $p_o^* - p_u^* < c_o - c_u$, where $p_o^* - p_u^*$ is as in (i), then rearrange to obtain $p_u^* - c_u > p_o^* - c_o$.

(iii) This follows immediately from (ii): gross profits are determined by squared markups, and $p_u^* - c_u > p_o^* - c_o$ for all parameter values. \square

Proof of Proposition 13:

(i) For a "pure official equilibrium" to exist, we must have:

$\Pi_o(\mu = 0, n; \cdot) = C_E$. Call the n that solves this equation n_o . (Such an n_o will always exist, as I assumed viability of the industry, i.e. $\Pi_o(\mu = 0, n = 1; \cdot) > C_E$.) Then, evaluate the underground firm's net profit at $(\mu = 0, n_o)$; if $\Pi_u(\mu = 0, n_o; \cdot) - \alpha F(\mu = 0) \leq 0$, there exists a "pure official equilibrium", with $(\mu^*, n^*) = (0, n_o)$.

For a "pure underground equilibrium" to exist, we must have:

$\Pi_u(\mu = 1, n; \cdot) - \alpha F(\mu = 1) = 0$. Call the n that solves this equation n_u . Then, evaluate the official firm's net profit at $(\mu = 1, n_u)$; if $\Pi_o(\mu = 1, n_u) - C_E \leq 0$, there exists a "pure underground equilibrium", with $(\mu^*, n^*) = (1, n_u)$.

Now, suppose that neither a "pure official equilibrium" nor a "pure underground equilibrium" exist, i.e. $\Pi_u(\mu = 0, n_o; \cdot) - \alpha F(\mu = 0) > 0$ and $\Pi_o(\mu = 1, n_u) - C_E > 0$. (If no $n_u \geq 1$ exists that solves $\Pi_u(\mu = 1, n; \cdot) - \alpha F(\mu = 1) = 0$, assume that $\Pi_o(\mu = 1, n) - C_E > 0$

holds for $n = 1$, which implies strengthening the viability assumption.) Then, note that Π_o is strictly decreasing in n ; thus, there must be an $n > n_u$, call it n_1 , that solves $\Pi_o(\mu = 1, n) - C_E = 0$. Since Π_u is decreasing in n as well, and $n_1 > n_u$, an underground firm's profits evaluated at $(\mu = 1, n_1)$ must be negative, i.e. $\Pi_u(\mu = 1, n_1; \cdot) - \alpha F(\mu = 1) < 0$.

Next, consider all pairs (μ, n) that set official firms' net profits equal to zero. Since Π_o is strictly and continuously decreasing in both μ and n , there will be a unique n for each value of $\mu \in [0, 1]$ such that official firms' profits are exactly equal to zero. In other words, $\Pi_o(\mu, n) - C_E = 0$ defines an implicit function $n(\mu)$ that is continuously decreasing in μ .

Likewise, Π_u is continuous in $(\mu, n(\mu))$ as defined above, and will therefore take on any value between $\Pi_u(\mu = 0, n_o; \cdot)$ and $\Pi_u(\mu = 1, n_1; \cdot)$ as we let μ run from 0 to 1. Recall that at $(\mu = 0, n_o)$, the underground firm's gross profits are strictly larger than $\alpha F(\mu = 0)$, while at $(\mu = 1, n_1)$, the underground firm's net profits are strictly smaller than $\alpha F(\mu = 1)$. By continuity of $\alpha F(\mu)$ in μ , $\alpha F(\mu)$ will take on all values between $\alpha F(\mu = 0)$ and $\alpha F(\mu = 1)$ as we let μ run from 0 to 1. Hence, there must be at least one $\mu \in (0, 1)$ such that $\Pi_u(\mu, n(\mu))$ and $\alpha F(\mu)$ intersect. Denote this value by μ^* , and denote $n(\mu^*)$ by n^* .

Then, at (μ^*, n^*) , we have $\Pi_u(\mu^*, n^*) - \alpha F(\mu^*) = 0$, and by construction of $n(\mu)$, we also have $\Pi_o(\mu^*, n^*) - C_E = 0$. Therefore, we found a pair (μ^*, n^*) that satisfies the equilibrium conditions for a coexistence equilibrium, which proves that if neither of the two pure equilibria exists, there must be at least one coexistence equilibrium.

(ii) Note first that the conditions for existence of a "pure official equilibrium" and of a "pure underground equilibrium" may be satisfied simultaneously, i.e. $\alpha F(\mu)$ and C_E may

be such that both $\Pi_u(\mu = 0, n_o; \cdot) - \alpha F(\mu = 0) \leq 0$ and $\Pi_o(\mu = 1, n_u) - C_E \leq 0$ hold.

Moreover, no matter if the two pure equilibria exist or not (or only one of them exists), there is nothing that prevents $\Pi_u(\mu, n(\mu))$ as defined above from intersecting more than once with $\alpha F(\mu)$ as μ runs from 0 to 1. To see this, recall that I have not imposed any restrictions on the shape of $\alpha F(\mu)$ (other than continuity in μ); now, while $\Pi_u(\mu, n(\mu))$ can be shown to be monotonically increasing in μ , $\alpha F(\mu)$ need not be monotonic in μ , thus allowing for more than one intersection with $\Pi_u(\mu, n(\mu))$. In fact, the number of coexistence equilibria can be arbitrarily large: define $\alpha F(\mu)$ to be exactly equal to $\Pi_u(\mu, n(\mu))$ for some or all μ on the interval $[0, 1]$ to obtain infinitely many coexistence equilibria. \square

Proof of Proposition 14:

Let (μ^*, n^*) be a coexistence equilibrium, so that $\Pi_o(\mu^*, n^*; \gamma) - C_E = 0$ and $\Pi_u(\mu^*, n^*; \gamma) - \alpha F(\mu^*) = 0$ both hold. Then, we can take the total differential of both equations at solution (μ^*, n^*) to have:

$$\begin{aligned} d\{\Pi_o(\cdot) - C_E\} &= \frac{\partial \Pi_o}{\partial \gamma} d\gamma + \frac{\partial \Pi_o}{\partial \mu} d\mu + \frac{\partial \Pi_o}{\partial n} dn \\ d\{\Pi_u(\cdot) - \alpha F(\mu^*)\} &= \frac{\partial \Pi_u}{\partial \gamma} d\gamma + \left(\frac{\partial \Pi_u}{\partial \mu} - \frac{\partial \alpha F}{\partial \mu} \right) d\mu + \frac{\partial \Pi_u}{\partial n} dn \end{aligned}$$

Note that if $d\mu$ and dn represent adjustments of μ and n to a new equilibrium, following a change in γ , we must have

$$\begin{aligned} d\{\Pi_o(\cdot) - C_E\} &= 0 \text{ and} \\ d\{\Pi_u(\cdot) - \alpha F(\mu^*)\} &= 0 \end{aligned}$$

These two equations allow us to solve for $\frac{d\mu^*}{d\gamma}$, i.e. the change in the equilibrium share of

firms in the underground economy relative to the change in the competition parameter γ , which yields:

$$\frac{\partial \mu^*(\cdot)}{\partial \gamma} = \frac{\frac{\partial \Pi_u}{\partial n} \frac{\partial \Pi_o}{\partial \gamma} - \frac{\partial \Pi_o}{\partial n} \frac{\partial \Pi_u}{\partial \gamma}}{\left(\frac{\partial \Pi_u}{\partial \mu} - \frac{\partial \Pi_o}{\partial \mu} \right) \frac{\partial \Pi_o}{\partial n} - \frac{\partial \Pi_u}{\partial n} \frac{\partial \Pi_o}{\partial \mu}}$$

Now, to evaluate the sign of this expression, first note that the following inequalities apply: both the official and the underground firm's gross profits are decreasing in μ , n and γ ; the underground firm's profits drop faster than the official firm's profits when μ or n increases, while the opposite is true for an increase in γ :

$$\frac{\partial \Pi_o(\cdot)}{\partial \mu} < 0, \frac{\partial \Pi_o(\cdot)}{\partial n} < 0, \frac{\partial \Pi_o(\cdot)}{\partial \gamma} < 0$$

$$\frac{\partial \Pi_o(\cdot)}{\partial \mu} - \frac{\partial \Pi_u(\cdot)}{\partial \mu} = 2\frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) \frac{\partial p_u^*}{\partial \mu} [(p_o^* - c_o) - (p_u^* - c_u)] > 0$$

because $2\frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) > 0$, $\frac{\partial p_u^*}{\partial \mu} = \frac{\partial p_o^*}{\partial \mu} < 0$, and $(p_o^* - c_o) - (p_u^* - c_u) < 0$ by Lemma 12.

$$\begin{aligned} \frac{\partial \Pi_o(\cdot)}{\partial n} - \frac{\partial \Pi_u(\cdot)}{\partial n} &= 2\frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) \left[(p_o^* - c_o) \frac{\partial p_o^*}{\partial n} - (p_u^* - c_u) \frac{\partial p_u^*}{\partial n} \right] + \\ &+ \left(-\frac{1}{n^2} \right) \left(1 + \gamma - \frac{2\gamma}{n} \right) [(p_o^* - c_o)^2 - (p_u^* - c_u)^2] > 0 \end{aligned}$$

because $2\frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) > 0$, $(p_o^* - c_o) \frac{\partial p_o^*}{\partial n} - (p_u^* - c_u) \frac{\partial p_u^*}{\partial n} > 0$ by Lemma 12 and $\frac{\partial p_o^*}{\partial n} - \frac{\partial p_u^*}{\partial n} > 0$,

$\left(-\frac{1}{n^2} \right) \left(1 + \gamma - \frac{2\gamma}{n} \right) < 0$, and $(p_o^* - c_o)^2 - (p_u^* - c_u)^2 < 0$ by Lemma 12.

$$\begin{aligned} \frac{\partial \Pi_o(\cdot)}{\partial \gamma} - \frac{\partial \Pi_u(\cdot)}{\partial \gamma} &= 2 \frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) \left[(p_o^* - c_o) \frac{\partial p_o^*}{\partial \gamma} - (p_u^* - c_u) \frac{\partial p_u^*}{\partial \gamma} \right] + \\ &+ \frac{1}{n} \left(1 - \frac{1}{n} \right) [(p_o^* - c_o)^2 - (p_u^* - c_u)^2] < 0 \end{aligned}$$

because $2 \frac{1}{n} \left(1 + \gamma - \frac{\gamma}{n} \right) \left[(p_o^* - c_o) \frac{\partial p_o^*}{\partial \gamma} - (p_u^* - c_u) \frac{\partial p_u^*}{\partial \gamma} \right] < \frac{1}{n} \left(1 - \frac{1}{n} \right) [(p_u^* - c_u)^2 - (p_o^* - c_o)^2]$ since $\frac{\partial \Pi_u(\cdot)}{\partial \gamma} < 0$ and $\frac{\partial \Pi_o(\cdot)}{\partial \gamma} < 0$, $\frac{1}{n} \left(1 - \frac{1}{n} \right) > 0$, and $(p_o^* - c_o)^2 - (p_u^* - c_u)^2 < 0$ by Lemma 12.

Given these inequalities, we can conclude that the numerator of $\frac{\partial \mu^*(\cdot)}{\partial \gamma}$ will be strictly positive. Thus, the sign of $\frac{\partial \mu^*(\cdot)}{\partial \gamma}$ is determined by the sign of its denominator. Now, if the denominator is positive, this is equivalent to having:

$$-\frac{(\partial \Pi_u / \partial \mu - \partial \alpha F / \partial \mu)}{\partial \Pi_u / \partial n} < -\frac{\partial \Pi_o / \partial \mu}{\partial \Pi_o / \partial n}$$

The left-hand side of this inequality is equivalent to the $[dn/d\mu]_u < 0$ that solves $d\{\Pi_u(\cdot) - \alpha F(\mu^*)\} = 0$ when γ is kept constant, i.e., $[dn/d\mu]_u$ identifies the line in the (μ, n) space along which the underground firm's profits are unchanged. The right-hand side of the inequality is the corresponding $[dn/d\mu]_o < 0$ that solves $d\{\Pi_o(\cdot) - C_E\} = 0$ when γ is kept constant.

Note that the equilibrium pair (μ^*, n^*) is the intersection of the two lines $[dn/d\mu]_u$ and $[dn/d\mu]_o$. Now, if $[dn/d\mu]_u < [dn/d\mu]_o$, then this implies that anywhere on the $[dn/d\mu]_o$ line to the right of μ^* , i.e. where $\mu > \mu^*$, the underground firms would make negative profits, thus inducing them to leave the underground sector until μ is back to its equilibrium value. (If the underground firms were instead to make positive profits, further entry into the underground sector would occur, until the industry settles at a new, pure underground, equilibrium).

Likewise, anywhere on the $[dn/d\mu]_o$ line to the left of μ^* , i.e. where $\mu < \mu^*$, the underground firms would make positive profits, thus inducing more firms to enter the underground sector until μ is back to its equilibrium value.

In other words, if $[dn/d\mu]_u < [dn/d\mu]_o$, this means that the coexistence equilibrium (μ^*, n^*) is stable (the industry will revert to this equilibrium after a small perturbation, rather than moving to an entirely different equilibrium); then, the denominator of $\frac{\partial \mu^*(\cdot)}{\partial \gamma}$ will be strictly positive as well, and this implies that $\frac{\partial \mu^*(\cdot)}{\partial \gamma} > 0$, as stated in the Proposition. \square

Appendix B: Descriptive Statistics and First-Stage Regressions

Let us first describe the remaining variables that enter the regression analysis in more detail:

(a) Low corruption

Source: GCR. The underlying question is: "Irregular, additional payments connected with import and export permits, business licenses, exchange controls, tax assessments, police protection, or loan applications are very rare (1=strongly disagree, 7=strongly agree)". This variable is a measure for good governance, and should therefore capture the rigor of enforcement represented by the expected fine $\alpha F(\mu)$ in the model.

(b) Labor regulation

Source: "Employment Laws Index" from the World Bank's "Labor Regulation" database. The employment laws index covers three areas: alternative employment contracts, conditions of employment, and employment protection. The employment laws index takes

values between 0 and 3, where the sub-indices on alternative employment contracts, conditions of employment, and employment protection each take values between 0 and 1. Higher values mean more regulation. This variable, an index of labor market flexibility, should measure the extent of government regulations that may increase the cost of production for those firms that comply with them, and is thus a proxy of the "non-tax cost advantage" of firms that fail to comply with them (because they operate in the underground economy). In terms of the model, this variable should partly reflect $c_o - c_u$.

(c) Tax rates

I will use tax rates for four different types of taxes: payroll tax, income tax, corporate tax, and VAT.²¹

(c1) Payroll tax rates (sum of employer and employee), % of earnings

Source: GCR (which quotes from the Social Security Administration's report "Social Security Programs Throughout the World"). This variable is a proxy for the cost advantage of hiring "unofficial" labor, and will therefore represent $c_o - c_u$ in terms of the model.

(c2) Highest marginal tax rate, individual rate (%)

Source: World Bank, "World Development Indicators". This variable should control for the taxation of income/profits earned by the self-employed and owners of small firms (not present in the model).

(c3) Highest marginal tax rate, corporate rate (%)

Source: World Bank, "World Development Indicators". This variable should control for

²¹Note that the tax *rate* is a somewhat incomplete measure of the actual tax burden, as it does not account for differences in the definition of the tax *base* across countries.

the taxation of income/profits earned by corporations (not present in the model).

(c4) Value added tax rate

Source: World Bank, "World Development Indicators". This variable should capture the tax burden on final sales (not incorporated in the model).

(d) New Business

Source: World Bank, "Doing Business" database. The data series measures the cost of starting a new business in percent of GNI per capita. This variable should capture the presence of barriers to entry, in particular administrative barriers etc., thus representing C_E (red-tape barriers to entry) in the model.

(e) Local Competition

Source: GCR. The underlying question is: "Competition in local markets is intense and market shares fluctuate constantly (1=strongly disagree, 7=strongly agree)". This variable should capture the intensity of competition between firms in the local industry, thus representing γ in terms of the model.

Table 1A presents summary statistics of the stacked data underlying the regression analysis, while Table 1B reports the corresponding correlation coefficients.

Table 1A: *The Data - Summary Statistics*

	stacked data			
	Min.	Max.	Mean	Std. Dev.
Underground Economy:				
1. currency-demand approach (UE1)	7.8	66.1	23.22	12.83
2. electricity approach (UE2)	7.5	68.0	22.2	13.08
Low corruption (LCorr)	1.37	6.91	4.80	1.55
Labor regulation (LR)	0.81	2.35	1.53	0.45
Payroll tax rate (PT)	0.00	61.00	27.44	15.80
Income tax rate (IT)	0.00	60.00	31.41	13.80
Corporate tax rate (CT)	15.00	55.00	31.63	6.63
VAT (VAT)	0.00	31.00	15.78	6.52
Cost of Start-up (CS)	0.4	269	19.52	40.05
Local Competition (LComp)	1.66	6.50	4.38	1.31

For an explanation of the abbreviations used in Table 1B, see Table 1A above.

Table 1B: *Correlations*

	UE2	LCorr	LR	PT	IT	CT	VAT	CS	LComp
UE1	0.98	-0.47	0.55	0.04	-0.33	-0.28	0.07	0.07	-0.17
UE2		-0.55	0.52	-0.02	-0.28	-0.27	0.00	0.05	-0.23
LCorr			-0.52	-0.17	0.37	-0.03	0.05	-0.37	0.52
LR				0.49	-0.05	0.02	0.21	0.06	-0.16
PT					0.27	0.26	0.53	-0.24	0.01
IT						0.23	0.40	-0.19	0.41
CT							0.09	0.02	0.05
VAT								-0.20	-0.06
CS									-0.13

Table 2 shows the results of the first-stage regressions, where each country's tax rates are predicted by the corresponding tax rate of a neighboring country (plus a country-fixed effect).

Table 2: *First-Stage Regression Results for Tax Rates*

	Dependent Variable			
	payroll tax	income tax	corporate tax	VAT
Sample: 1995 - 2000 (6 years)				
payroll tax rate*	0.0577 (0.38)			
income tax rate*		0.75 (0.00)		
corporate tax rate*			0.09 (0.17)	
VAT rate*				0.01 (0.53)
country-years	262	264	277	272
# of country dummies	52	56	56	58
adjusted R^2	0.91	0.65	0.53	0.95

* = Neighboring country's

Appendix C: Robustness Checks

Table 4 shows the results of performing the same regressions as in Section 3, but including four regional dummies, where each country was assigned to either the OECD, the Central and Eastern European (CEE) countries, Latin America, or Asia.

APPENDIX C: ROBUSTNESS CHECKS

Table 4: *OLS Regression including regional dummies*

Sample: 1995 -2000 (6 years)		
OLS I: Dep. Var.: "Underground Economy" (currency-demand approach)		
OLS II: Dep. Var.: "Underground Economy" (electricity approach)		
	OLS I	OLS II
low corruption	-2.49 (0.0624)	-4.76 (0.0000)
labor regulation	11.26 (0.0000)	11.23 (0.0000)
payroll tax % IV	-0.23 (0.0068)	-0.31 (0.0000)
income tax % IV	-0.16 (0.1845)	0.04 (0.6003)
corporate tax % IV	0.01 (0.9596)	-0.25 (0.0442)
VAT % IV	0.51 (0.0090)	0.25 (0.0586)
Cost of Start-up (% GNI)	-0.06 (0.0377)	-0.07 (0.0000)
Competition	2.59 (0.0405)	1.45 (0.0521)
OECD	6.58 (0.5301)	33.14 (0.0000)
CEE	11.83 (0.2542)	33.43 (0.0000)
Latin America	16.30 (0.0900)	37.76 (0.0000)
Asia	14.40 (0.0833)	35.41 (0.0000)
country-years	102	248
Adjusted R^2	0.51	0.53

Table 5a shows the first-stage regression of "competition" on "fuel, ores, and exports", "average tariff rate", "effectiveness of antitrust policies", the log of popu

and country dummies.

Table 5a: *First-Stage Regression Results for "competition"*

Sample: 1995 - 2000 (6 years)	
Dependent Variable: Competition	
fuel, ores, and metal exports	-0.02 (0.0872)
average tariff rate	-0.03 (0.0019)
antitrust policies	1.30 (0.0000)
log of population	13.78 (0.0000)
country-years	303
# of country dummies	56
adjusted R^2	0.68

Table 5b presents the results of performing the same regressions as in Section 3, but instrumenting for "competition" by the regressors of Table 5a.

Table 5b: *OLS Regressions - "competition" instrumented*

Sample: 1995 -2000 (6 years)		
OLS I: Dep. Var.: "Underground Economy" (currency-demand approach)		
OLS II: Dep. Var.: "Underground Economy" (electricity approach)		
	OLS I	OLS II
constant	20.23 (0.0223)	38.51 (0.0000)
low corruption	-2.96 (0.0093)	-5.00 (0.0000)
labor regulation	12.93 (0.0000)	11.83 (0.0000)
payroll tax % IV	-0.25 (0.0039)	-0.33 (0.0000)
income tax % IV	-0.19 (0.1363)	-0.01 (0.8784)
corporate tax % IV	-0.26 (0.1632)	0.34 (0.0029)
VAT % IV	0.46 (0.0206)	0.25 (0.0438)
Cost of Start-up (% GNI)	-0.05 (0.0830)	-0.07 (0.0000)
Competition IV	2.59 (0.0600)	1.58 (0.0892)
country-years	97	239
Adjusted R^2	0.41	0.50

1. *...*
2. *...*

3. *...*
4. *...*

5. *...*

6. *...*

7. *...*

8. *...*

9. *...*

10. *...*

11. *...*

12. *...*

13. *...*

14. *...*

15. *...*

16. *...*

