

A Search-Theoretic Model of the Retail Market for Illicit Drugs*

Manolis Galenianos
Pennsylvania State University

Rosalie Liccardo Pacula
RAND Corporation

Nicola Persico
Northwestern University

November 2, 2011

*We thank Imran Rasul and three referees for comments that have improved the paper. We thank Emel Yildirim for excellent research assistance. We also thank, without implicating, Peter Reuter who gave the impulse for writing this paper. We thank Christian Ben Lakhdar, Jon Caulkins, Ric Curtis, Boyan Jovanovic, Beau Kilmer, Rasmus Lentz, Jin Li, Iouri Manovskii, Jeffrey Miron, Theodore Papageorgiou, Chris Pissarides, Larry Samuelson, Tom Sargent, Jose Scheinkman, Robert Shimer, Gianluca Violante, Travis Wendel, Randy Wright and Antonis Zagaris for useful discussions. We thank participants of the NBER's 2007 Economics & Crime Meetings and 2008 Summer Institute, the University of Maryland's 2007 "Economics and Crime" Conference, the 2007 SED, Summer Meetings of the Econometric Society, and Midwest Macro Meetings, the Penn Search and Matching Workshop, as well as participants of a number of seminars. Galenianos and Persico thank the National Science Foundation for financial support (grant SES-0922215). Pacula's time on this project was supported by a grant from the National Institute on Drug Abuse to the RAND Corporation (Grant R01 DA019993-01A1). Persico's work on this project was mostly carried out while in the employment of New York University.

Abstract

A search-theoretic model of the retail market for illegal drugs is developed. Trade occurs in bilateral, potentially long-lived matches between sellers and buyers. Buyers incur search costs when experimenting with a new seller. Moral hazard is present because buyers learn purity only after a trade is made. This model is consistent with some new stylized facts about the drugs market, and it is informative for policy design. The effectiveness of different enforcement strategies is evaluated, including some novel ones which leverage the moral hazard present in the market.

JEL codes: K42; J64

Keywords: Search theory; drugs; crime

1 Introduction

The market for narcotics is the cause of many social ills in the United States. The trade in illicit drugs gives rise to an underground economy that generates addiction, crime, and violence. In less affluent and minority communities, the drug economy crowds out the incentives to join the formal sector and it raises incarceration rates. In an effort to counter these trends, massive amounts of resources are devoted to interfering with the drugs market—the so-called “war on drugs”. This massive intervention takes place under a conception of the drugs market as a Walrasian market: a centralized market with the usual demand and supply curves, and a market-clearing price. While the Walrasian paradigm provides many important insights, we show that it fails to capture a number of empirical stylized facts about the retail drugs market. We propose another model, one of search with moral hazard, which does. The aim of this exercise is not merely descriptive; the model gives a more nuanced view of the effectiveness of some current policy interventions, and it also suggests new channels for effectively interfering with the retail market.

The model is one of repeated trade with unobservable quality. The focus of the analysis is to determine what level of quality will be traded for a given amount of money, that is, the *affordability* of (high quality) drugs in equilibrium. Formally, we build on the standard search model of Burdett and Mortensen (1998). Searching for sellers is costly. A seller always offers the same quality to a given buyer. Over time, a buyer who starts off unmatched searches until he finds a suitably high-quality seller, at which point he matches with that seller. The match persists until either (a) it is permanently broken up (for example, the seller goes to jail); or (b) during an occasional temporary disruption of the match (maybe the regular seller cannot be located that day) the buyer samples a different seller who happens to sell better quality, in which case he switches to the new seller. We modify the standard Burdett-Mortensen setup by assuming that buyers can only determine the quality of drugs *after* the trade is consummated. Introducing this moral hazard into the standard search model generates novel predictions about “price dispersion,” especially the presence of a fraction of sellers who sell zero quality.

The predictions from our model are consistent with a number of new stylized facts, which we document. The first is the presence of *rip-offs*, transactions in which the buyer is sold

essentially zero-purity drugs at a price that is not distinguishable from that of “regular” transactions. The second fact is the presence of long-term relationships between buyer and seller. The third fact is the presence of considerable dispersion in the price/quality ratio. These stylized facts are not accounted for by a Walrasian model.

This model can help us evaluate existing policy in a more nuanced way. The conventional view is rather generic: tougher penalties and more law enforcement, at any level of the supply chain, should help reduce the affordability of drugs. In fact, there is little evidence that recent efforts to increase penalties and law enforcement have measurably reduced the availability of drugs.¹ Our model offers a more nuanced view: different enforcement instruments can impact the retail affordability of drugs in complex and sometimes counterintuitive ways. For example, to the extent that police enforcement makes it more risky to search for new sellers, the long-term relationship between buyers and sellers is strengthened, which in turn alleviates moral hazard and improves the equilibrium price/quality ratio. Thus the market price need not be related to the intensity of interdiction in the expected way. Such findings highlight the need for an accurate model of market structure in order to evaluate existing policy.

At a somewhat more speculative level, the analysis suggests alternative channels to suppress the market. If it’s true that the market is undermined by moral hazard, and we think this paper makes a strong case that it is, then economic theory suggests *leveraging the moral hazard*, i.e., inducing sellers to *dilute more*. We will suggest a sentencing scheme that can help achieve this goal and simultaneously decrease the number of incarcerated sellers. The scheme works by reducing the sentence of sellers who are caught selling diluted drugs.

Much of the previous literature on illicit drugs markets has focused on modeling the *demand* for illicit drugs, discussing the role of harmful addiction, rationality, and discounting (Grossman and Chaloupka, 1998; Becker and Murphy, 1988; Schelling 1984; Stigler and Becker, 1977). Formal theoretical models of the *market structure* are very sparse and tied to traditional economic assumptions of perfect information and/or a centralized market—see Bushway and Reuter (2008) for a review article. Within this framework, all types of enforcement at all levels of the supply chain are generally lumped together and modeled as a “cost of

¹The price per pure gram of cocaine and heroin have declined substantially during the periods when budgets on law enforcement rose and penalties increased (Caulkins *et al.*, 2004).

doing business” for the dealer.² This framework abstracts from the defining features of illicit markets: non-contractibility and search costs, and so it ignores two relevant avenues through which law enforcement can influence the market: by increasing search time and influencing the distribution of purity in the market.³

Search models are somewhat related to switching cost models.⁴ In those models, firms initially compete intensely for market share and then, after switching costs take hold, they behave more monopolistically. None of these models, to our knowledge, features moral hazard,⁵ and so the prevalence of rip-offs is not easily interpreted through the lens of switching costs alone. That said, switching cost models have the interesting feature that a buyer should get progressively worse deal from his seller. Although we have at present no empirical evidence regarding this phenomenon, switching cost models may prove useful in the study of the evolution of the terms of trade during a buyer-seller relationship.⁶

A number of papers in the monetary search literature have dealt with the issue of decentralized trade under asymmetric information, e.g. Williamson and Wright (1994), Trejos (1999) and Berentsen and Rocheteau (2004). In all three papers, the buyers are (potentially) unable to assay the quality of the transacted good which is chosen strategically by the sellers. While similar to our model in many respects, these papers only consider one-off transactions between two agents; in contrast, we focus on the interplay between asymmetric information and repeated interactions. Crime has been introduced in search models to examine the interaction between the potential for crime opportunities that individuals face and their labor market outcomes, as in Burdett *et al.* (2003), Huang *et al.* (2004), and Engelhardt *et al.* (2008).

²Within this framework, Becker *et al.* (2006) argue that the market should be regulated by taxing rather than interdicting. The basic argument is that taxes could be levied at low administrative cost, while interdiction is costly to enact and to evade. Of course, if taxes on “legalized” drug are high then there would be “illegal” (i.e., tax-evading) drug sellers, and so our analysis would still apply to that segment of the market.

³Reuter and Caulkins (2004) represents a commendable exception, in that they document the large price and quality dispersion in the drugs market, and they informally conjecture that it may be connected to search frictions and/or moral hazard. Their paper does not develop a formal model, however.

⁴Beggs and Klemperer (1992), Klemperer (1987, 1995)

⁵Farrell and Shapiro (1989) come closest; they assume that quality is observable but not contractible.

⁶Our model is also tangentially related to the IO literature that studies a firm’s quality decision in a market for experience goods. Since these papers examine the markets for legal commodities, matching frictions play a relatively minor role. In contrast, in the market that we are looking at the frictions and turnover of both buyers and sellers are very important. A notable exception is Gale and Rosenthal (1994) where buyers have to pay a cost before finding a high-quality seller.

The rest of the paper is organized as follows. In the next section we present the stylized facts that motivate our theoretical analysis. In Section 3 we introduce the model and define the equilibrium notion. The equilibrium is characterized in Section 4. Section 5 obtains some testable implications and compares them with available data. Section 6 presents our results concerning the effect of existing enforcement policies, and analyzes the effect of alternative policies. Section 7 extends the basic model to include endogenous demand. Section 8 concludes.

2 Stylized Facts

We concentrate our attention on the heroin, crack cocaine and powder cocaine markets. Our information regarding drug markets and how buyers and sellers transact comes from two primary data sources: the System to Retrieve Information from Drug Evidence (STRIDE) database and the Arrestee Drug Abuse Monitoring (ADAM) Program. We use information available in the 1981-2003 STRIDE which include the type of drug obtained (heroin, cocaine, marijuana, methamphetamines...), method of acquisition (undercover purchase or seizure), price (in the case of a purchase), city and date of acquisition, quantity, as well as the purity level of the drug.⁷ The unit of analysis is the single transaction. The STRIDE data come from police informants and undercover agents working for a variety of law enforcement agencies. The reliability of the STRIDE data set is discussed in Appendix B. For our purposes, a critical feature of the data is that it is collected by police agencies and thus, probably, more representative of first-time transactions than of the kind of long-term buyer-seller interaction our model predicts.⁸ Fortunately, our model also yields predictions for the distribution of first-time transactions; it is this distribution which we compare to the data. We restrict our analysis to street-level STRIDE transactions, which we define as those worth less than \$100 in 1983 dollars. It is for these transactions that the moral hazard problem, which is central to the model, is most likely to be important.⁹ We will focus on *pure quantity*, defined as the product of (raw) quantity times purity, as our measure of the value of a trade to the buyer.

⁷The latter is determined through chemical analysis in a DEA laboratory

⁸Unfortunately, the nature of the relationship between buyer and seller is not disclosed in the STRIDE extract of the data made available to us.

⁹For large transactions involving many thousands of dollars, it is likely that methods to assay the drugs would be available to the buyer, and so the moral hazard problem would be much reduced.

The ADAM data set is collected quarterly from interviews with persons arrested or booked on local and state charges in various ADAM metropolitan areas in the United States. The sample contains demographic data on each arrestee, data on alcohol and drug use, abuse and dependence, and the drug acquisition data covering five most commonly used illicit drugs. Information collected includes number of times drugs were purchased and consumed in the past 30 days, number of drug dealers they transacted with, whether they last purchased from their regular dealer, difficulties experienced in locating a dealer or buying the drug, and the price paid for the specific quantity purchased. See Appendix B for more information on the ADAM data.

The following table, which is based on STRIDE data, shows that retail transactions for illegal drugs are subject to moral hazard: that is, the seller can covertly dilute (“cut”) the product, and this dilution is largely unobservable to buyers until after they consume. The table documents an extreme instance of the moral hazard—the *rip-off*, a transaction in which the buyer is sold essentially zero-purity drugs. We label as rip-offs those trades that yield a pure quantity which is less than 2% of the average pure quantity traded. A significant fraction of “street-level” transactions are seen to be total rip-offs. Most important, the price paid in a rip-off is not appreciably different from that of non-rip-off transaction, suggesting that buyers cannot observe dilution.^{10,11}

¹⁰Even the relatively high incidence of rip-offs found in Table 1 may underestimate the extent of cheating in this market. This would be the case if the STRIDE data included purchases from trusted, or regular sellers, because these sellers are presumably less likely to cheat their customers.

¹¹The practice of selling drugs in branded bags (“dope stamps”) is further corroborating evidence of a quality problem in the illegal drugs market. Dope stamps could be boasts of quality (“America’s Choice,” “Dynamite”), status brands (“Dom Perignon,” “Gucci”), and even corporate names (“Exxon”). The purported effect of a dope stamp is quality certification. However, because the stamps can be faked by “unscrupulous” competitors, the certification value of a dope stamp is limited and often very short-lived (a couple of days, often). Not very much is known about the phenomenon of dope stamps: Wendel and Curtis (2000), for example, report in their interesting study that dope stamps are apparently limited to heroin sales in or around New York City—exactly why it is not clear. What seems clear, however, is that dope stamps did not solve the quality certification problem.

Drug	average pure quantity in grams	percentage of all trades that are rip-offs $\leq 0.003g$ for heroin $\leq 0.01g$ for crack cocaine $\leq 0.01g$ for powder cocaine	average price of rip-offs (std. dev. of price)	average price of non rip-offs (std. dev. of price)
Heroin	0.16	8.2%	\$50.1 (22.5)	\$56.9 (20.6)
Crack Cocaine	0.46	7.9%	\$31.5 (21.3)	\$37.6 (24.6)
Powder Cocaine	0.64	5.3%	\$34.4 (21.5)	\$53.2 (25.8)

Table 1: Pure quantity of trades with value \leq \$100 in 1983 dollars.¹²

Long-term relationships are the second basic fact that we document.¹³ The next table, compiled from the ADAM data set, provides (buyer-reported) evidence of a large amount of repeat business. Each buyer is asked to report whether the last person from whom he purchased drugs was a regular, occasional, or new supplier. Overall there is a lot of repeat business. Table 2 shows that for heroin, for example, more than 58% of users obtained their last purchase from their regular supplier. The presence of repeat business is consistent with the equilibrium of our model.

Last supplier	Heroin		Crack Cocaine		Powder Cocaine	
	Frequent	Casual	Frequent	Casual	Frequent	Casual
Regular	76%	58%	62%	43%	79%	57%
Occasional	18%	26%	27%	35%	15%	28%
New	6%	16%	10%	22%	6%	15%

Table 2: Repeated transactions.¹⁴

¹²Prices computed in 1983 dollars. The number of observations is 12,721 for heroin, 16,202 for crack cocaine, and 5,362 for powder cocaine.

¹³Not all sellers need have repeat business. The ethnographic literature also reports of sellers who specialize into selling rip-offs. In our model, these sellers will be called “opportunistic sellers” and will have no repeat business. Hamid (1992, p. 342) refers to these sellers as “zoomers,” a street expression due to the practice of selling bogus drugs and then disappearing.

¹⁴For each drug, these are the male respondents who reported consuming that drug at least once in the previous 30 days. Frequent consumers are those who report using the drugs more than 20 times in a month.

Table 2 reveals further detail. It indicates that more frequent consumers appear to be more loyal to their regular suppliers. A search model such as ours would yield a similar prediction if it featured heterogeneity in the frequency of consumption.

The third basic fact is the presence of considerable dispersion in how much pure drugs a given amount of money can buy. In Appendix C we document this dispersion and offer evidence that is not an artefact of aggregation across time and space, by showing that a large dispersion persists after taking out fixed effects for time and place of the transaction.¹⁵ Since in a Walrasian market we would expect the “law of one price” to hold, we view this large quality dispersion as evidence in favor of a model with search frictions, such as the one presented in this paper.

3 Model and Equilibrium Definition

Time runs continuously, the horizon is infinite, and the future is discounted at rate r . There is a continuum of buyers (or customers) of measure β . For now, we treat β as exogenous and we endogenize it in Section 7. There is a continuum of sellers (or suppliers) of measure σ .¹⁶ A free entry condition with entry cost K determines the mass of sellers σ who participate in the market. Buyers want to trade with sellers.

Each buyer gets the urge/ability to consume at random times which arrive at Poisson rate α . When a consumer gets the urge/ability to consume, he takes a sum of money m and purchases whatever drugs he can. One way to think about this process is that addicts will, with Poisson rate α , be able to obtain m dollars, which they immediately use to purchase drugs. The available evidence from the ADAM data set suggests that these urges to consume are pretty frequent for many consumers. For instance, out of all ADAM respondents admitting to heroin use in the previous thirty days, almost 60% report buying heroin at least 28 times during the past month, and over a third report buying it multiple times in a single day.¹⁷

¹⁵We show that a large amount of dispersion persists even after we break down transactions by (non-pure) weight (although we do find that bigger transactions are associated with a higher pure-gram-per-dollar amount, which can be viewed as evidence of “quantity discounts.” We thank a referee for suggesting that we look into quantity discounts.

¹⁶Sellers in our model could be single pushers or criminal gangs.

¹⁷The sheer frequency of these purchases suggest that buyers don’t store drugs very much. This impression is corroborated by the very high correlation in the ADAM data between the number of purchases and the

For simplicity, m is exogenously given and is the same for all consumers. Empirically, we take m to be a relatively small sum (less than \$100 in 1983 dollars), because it is for these transactions that the moral hazard problem is most likely to be important.¹⁸

In return for m , the buyer receives q . We will refer to q as *quality*. q represents an aggregator of quantity and purity, and it captures the utility that the buyer receives from consuming. Empirically, we will proxy for quality by *pure quantity* defined as the product of (raw) quantity times purity.

While m is observed by both buyer and seller, the quality q fetched by m cannot be determined by the buyer at the time of the transaction. Quality q is chosen by the seller, through “cutting.” After the buyer consumes the good, the quality of the purchase is perfectly revealed. This ex-post knowledge affects the buyer’s decision of whether to *match* with the seller (see below). The seller pays c per unit of quality that he supplies to the buyers that visit him.^{19,20} The main assumption on sellers’ behavior is that, once they decide on the quality level that they offer a particular buyer, they commit to their decision forever. That is, a seller supplies the same quality to a particular buyer at all times and, as a result, the buyer knows the quality that he will receive from a particular seller once he has sampled from him.^{21,22}

The market is characterized by search frictions in the sense that there is no central market-

number of times users report consuming the drugs in a month (0.89 for heroin, 0.89 for crack and 0.82 for powder cocaine).

¹⁸We take m as exogenous. In reality, consumers—even addicts—have a choice of how much money to devote to drugs consumption. Such consumers presumably trade off their opportunity cost for money $O(m)$ against the quality $Q(m)$ that money can fetch. Our present analysis pins down $Q(m)$. One could then specify some function $O(m)$ and obtain the optimal (endogenous) m^* . After the optimal m^* is determined, our analysis applies directly.

¹⁹While we model the seller as having the ability to personally cut the drugs, an alternative interpretation of our formal model would be that sellers do not cut themselves, but rather can procure drugs of different purities from a wholesale “quality menu” (at wholesale prices that reflect purity, of course).

²⁰The cost c at which sellers procure pure drugs could be set by an upstream monopolist, or if the sellers were integrated with the monopolist, c would represent the shadow cost of capital for this monopoly.

²¹This assumption is less stringent than it might appear: Coles (2001) shows that commitment to a given quality level can arise as part of the equilibrium outcome of a broader model where sellers find it profitable to commit for reputational reasons. An alternative assumption is that a seller changes the quality that he offers to his customers at random intervals. We consider this extension in Galenianos *et al.* (2009) and show that the qualitative properties of our model do not change.

²²One might be concerned that it might be difficult for a seller to always provide the same quality to a customer when the wholesale purity becomes diluted. However, notice that we define quality as pure quantity, so a retailer faced with diluted wholesale drugs could keep up quality by simply selling more quantity to the buyer. Thus wholesale quality need not represent a “technological upper bound” on retail quality.

place where all agents can meet to trade. Rather, buyers and sellers have to trade bilaterally. In our model a buyer can be in either of two states: *matched*, which means that he has a regular supplier, or *unmatched*. An unmatched buyer has to search in the market at random, incurring utility cost of search, s . A matched buyer can still search at cost s , but he also has the option of visiting his regular supplier, which does not entail any cost. However, there is a probability γ that the regular supplier is unavailable, in which case even the matched buyer has to search at random and incur cost s . Since the matched buyer retains the option of going back to his “regular” seller in a future transaction, this event represents a temporary separation. Furthermore, a match between a buyer and a seller is exogenously destroyed at rate δ .

Search frictions are introduced in the model for good reason. In the highly decentralized market for illicit drugs, sellers cannot advertise their location or the quality of their products, so buyers have to expend resources trying to locate each other without attracting police attention. The temporary break-ups we assume in the model are observed in the data. Among the ADAM respondents who responded to detailed questions about heroin purchases, about a quarter report not being able to purchase heroin at some point in the past 30 days. In many cases, the causes they mention appear temporary in nature (e.g., “police activity,” and “no dealer available”). If these obstacles can prevent buyers from buying heroin, presumably they also drive buyers to temporarily experiment with new sellers. The permanent break-ups we assume in the model may be due to death or incarceration of either the buyer or the seller. Reuter, MacCoun and Murphy (1990) estimate that in 1988 in Washington DC the probability that a drug dealer became incarcerated was 22%. In addition, drug dealers faced a 1 in 70 annual risk of getting killed and 1 in 14 risk of serious injury.

The transition between the matched and unmatched state takes place after the trading is done. We now detail the transitions between the two states. An unmatched buyer decides whether to match with a seller after consuming his good. If this occurs, the seller becomes his regular supplier. A matched buyer who sampled a new seller because his regular supplier was unavailable decides whether to switch to the new seller or to return to his previous supplier. A match between a buyer and a seller is exogenously destroyed at rate δ and in this event the buyer becomes unmatched.

The focus of the analysis will be to determine what quality q will be offered by sellers in

equilibrium, in exchange for (any arbitrarily fixed) m . The ratio q/m represents the terms of trade, as it were—and can be thought to capture the *affordability* of drugs. We look for steady state equilibria, that is, equilibria in which the search strategy of buyers and the quality distribution sold by sellers are time-invariant. Of course, even in a steady state equilibrium a generic buyer searches while matched, and thus consumes progressively better-quality drugs. Random break-ups in the matches will set back this process.

3.1 Buyers’ Decision Problem

We first consider the buyers’ search problem, taking the sellers’ actions as exogenous. We will show that the optimal search strategy is to stop searching (and thus match) if and only if the quality offered by the current seller is above a threshold R . We also characterize R .

Let F denote an arbitrary distribution of qualities in the market with support in $[0, \bar{q}]$. The state variables for a buyer is whether he is matched and, if so, what is the quality that he receives from his regular supplier. Let $V(q)$ denote the value of being matched with a seller who offers quality q . Let \bar{V} denote the value function of a buyer who does not have a regular seller.

The value functions in flow terms are given by the following asset pricing equations (recall that r is the discount rate):

$$r \bar{V} = \alpha \left[-s + \int_0^{\bar{q}} (\tilde{q} + \max\{V(\tilde{q}) - \bar{V}, 0\}) dF(\tilde{q}) - m \right] \quad (1)$$

$$r V(q) = \alpha \left[(1 - \gamma) q + \gamma \int_0^{\bar{q}} (-s + \tilde{q} + \max\{V(\tilde{q}) - V(q), 0\}) dF(\tilde{q}) - m \right] + \delta (\bar{V} - V(q)). \quad (2)$$

The interpretation is as follows. Consider equation (1) first. At rate α the buyer gets the urge to consume. When this happens, he samples a seller at random and incurs the cost of search s . The instantaneous utility that he receives from consuming is a random draw from the distribution of qualities, F . After consuming, the buyer decides whether to keep this seller as his regular supplier, which yields a “capital gain” of $V(\tilde{q}) - \bar{V}$, or to remain unmatched, in which case there is no change in his value. In either case, he pays m to

the seller. Equation (2) is similar. Again, at rate α the buyer wants to consume. With probability $1 - \gamma$ his regular supplier is available and the buyer receives quality q . With probability γ the regular seller is unavailable and the buyer has to search in the market. As a result, he incurs cost s and he makes a random draw from F . The only difference from the previous case is that he compares the new seller with his regular supplier when deciding whether to stay with the new draw. Therefore the capital gain of switching to the new seller is $V(\tilde{q}) - V(q)$. Regardless of which seller he transacts with, the buyer pays m . At rate δ the match is destroyed and the buyer becomes unmatched leading to a capital loss of $\bar{V} - V(q)$. Note that $V(q)$ is strictly increasing in its argument. Thus there is a unique reservation value R . An unmatched buyer who samples a seller offering quality $q \geq R$ will choose to match with the current seller, while if $q < R$ he will remain unmatched. A matched buyer will switch suppliers if and only if the new seller offers a higher quality. We now characterize R as a function of the (still) exogenous distribution F by using the equilibrium condition $V(R) = \bar{V}$.

Lemma 1 *We have*

$$R = -s + \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \alpha (1 - \gamma) \int_R^{\bar{q}} \frac{1 - F(\tilde{q})}{r + \delta + \alpha \gamma (1 - F(\tilde{q}))} d\tilde{q}, \quad (3)$$

or $R = 0$ if the right-hand side is negative.

Proof. See Appendix A.1. ■

3.2 Sellers' Decision Problem

The seller's problem is to choose a level of quality that maximizes his steady-state level of profits. This formulation implies that sellers are arbitrarily patient. The steady state profits of a seller who chooses to offer quality level q are given by

$$\pi(q) = (m - c q) t(q) \quad (4)$$

The first terms is the seller's margin per sale (with a linear cost c of quality); $t(q)$ is the expected flow of transactions at the steady state, which will be characterized in Section 4.

3.3 Definition of Steady-State Equilibrium

Definition 1 *A steady-state equilibrium is a buyer reservation value R , a distribution F of sellers quality and a mass of sellers σ such that the following conditions hold:*

1. *Buyer optimization: $R = R(F)$ where $R(F)$ is defined in Lemma 1.*
2. *Seller optimization: seller's profits $\pi(q)$ equal $\bar{\pi}$ whenever q is offered in equilibrium, and otherwise $\pi(q) \leq \bar{\pi}$.*
3. *Free entry of sellers: the mass of sellers σ is such that the profit level equals the cost of entry, $\bar{\pi} = K$.*

4 Equilibrium Characterization

Equilibria in our model exist and are unique. Theorem 1 in Appendix A.2 establishes existence and uniqueness of an equilibrium. Depending on parameter values, the distribution of quality traded q may exhibit a mass point at zero—a feature of particular interest for us. Theorem 1 says that the equilibrium always exhibits a mass of sellers offering zero quality, if search costs s are sufficiently low. Intuitively, this is because when s is small, buyers are picky about which seller to match with, which in turn increases the sellers' incentives to cheat. Rather than offer high quality in order to get repeat business, more sellers will opt for the quick one-time profit and offer zero quality.

In the remainder of the paper we focus on equilibria with a (possibly very small) mass of sellers offering zero quality. Such equilibria are the empirically relevant ones because in our data we find a significant amount of zero-purity transactions. For completeness, in Appendix A.2 the equilibrium is characterized for any parameter configuration.

To characterize the quality distribution F , we use the fact that in equilibrium all qualities that are offered yield the same steady state profits. Offering a higher level of quality reduces the margin per transaction and increases the number of sales. For quality levels in the support of F , the two effects balance each other exactly.

A seller's sales come from two sources: his steady state number of "loyal" customers, and the new customers who sample once and may or may not match with the seller after consuming (if they do match, they are counted as loyal from then on). Denote by $t_L(q)$ and t_N , respectively, the flow of sales to loyal and new buyers by a seller offering quality q . The flow of sales per seller $t(q)$ is given by the sum of these two flows,

$$t(q) = t_N + t_L(q).$$

The number of loyal customers depends on F and is increasing in q both because a higher-quality seller has more competitors from whom to poach customers (higher inflow) and because there are fewer sellers that can poach his own customers (lower outflow).

We now characterize $t(q)$ and F .

Proposition 1 *For any m , the following properties hold in equilibrium:*

(i) *The steady state flow of transactions of a seller offering quality q is given by*

$$t(q) = \frac{\alpha\beta}{\sigma} \frac{\delta + \alpha \gamma (1 - F(R))}{\delta + \alpha (1 - F(R))} \left[1 + \frac{\alpha \delta (1 - \gamma)}{[\delta + \alpha \gamma (1 - F(q))]^2} \right], \quad \text{when } q \geq R, \quad (5)$$

$$t(q) = \frac{\alpha\beta}{\sigma} \frac{\delta + \alpha \gamma (1 - F(R))}{\delta + \alpha (1 - F(R))}, \quad \text{when } q < R. \quad (6)$$

(ii) *If q is offered by a seller then either $q = 0$ or $q \geq R$.*

(iii) *F has no mass point on the positive part of its support.*

(iv) *F exhibits quality dispersion.*

(v) *The positive part of the support of F is connected and is given by $[R, \bar{q}]$.*

(vi) $\bar{q} = \frac{m}{c} \cdot \frac{\alpha (1-\gamma)}{\delta + \alpha (1-\gamma)}$.

(vii) *On the positive part of its support, F is given by*

$$F(q) = 1 + \frac{\delta}{\alpha \gamma} - \frac{1}{\alpha \gamma} \sqrt{\alpha \delta (1 - \gamma)} \sqrt{\frac{(m/c) - q}{q}}.$$

(viii) *On the positive part of its support, F is concave if (and only if) $\frac{\alpha (1-\gamma)}{\delta + \alpha (1-\gamma)} \leq \frac{3}{4}$.*

Proof. See Appendix A.1. ■

Why is F non-degenerate, that is, why does quality dispersion arise in equilibrium? Suppose, by contradiction, that all sellers offered the same quality q^* . Then a seller who offered a slightly higher quality $q^* + \varepsilon$, would be able to retain all his current customers as well as poach every buyer that ever was ever temporarily matched with him. This would lead to a discrete increase in profits at a negligible cost (ε can be very small).

To interpret the condition in part (viii) of Proposition 1, it helps to rewrite it as $\bar{q} \leq \frac{3}{4} \frac{m}{c}$. The term \bar{q} measures the maximal quality offered on the market in equilibrium. The ratio m/c represents the quality that m could fetch in a competitive market without moral hazard, where sellers price at marginal cost. So the inequality identifies parameter constellations, those with relatively “large” separation rate δ , which gives rise to a quality distribution which is much inferior to the competitive one. Another interpretation is that the ratio measures how many transactions a buyer can get out of a long-term relationship before it breaks up.

To complete the characterization of the equilibrium, we need to determine the equilibrium mass of sellers σ^* . This is done in the next proposition.

Proposition 2 1. *The equilibrium quality distribution F does not depend on the buyer-seller ratio β/σ .*

2. *Profits per seller are multiplicative in β/σ .*

3. *Therefore, given β there exists a unique value σ^* that solves the free-entry condition $\bar{\pi} = K$.*

Proof. See Appendix A.1. ■

We view the “irrelevance”result in part 1. above as a convenient feature of our model, but not a fundamental one. It is convenient because, within our assumptions, it allows one to separate the question of entry from the rest of the analysis. It is not fundamental in the sense that changing some of the assumptions, e.g. those concerning the matching technology, would probably invalidate this stark result while preserving what we regard as the fundamental features of our model (price dispersion, incentives to dilute, etc.).

5 Testable Implications

The model has implications both for the cross-sectional distribution of qualities traded and for the time series of individual consumption. Now we look for these implications in the data.

The first implication is the presence of price dispersion. The model predicts that the same amount of money should, in equilibrium, fetch different amounts of pure drugs depending on the type of seller with whom the buyer is marched. The presence of quality dispersion is a very robust feature of the data and is documented in Appendix C. In that section we show that this dispersion persists after controlling for the observables that we have.²³ The degree of quality dispersion is also seen in Figure 2 below.

The second implication is the presence of a mass of transactions with zero quality, whose price is the same as that of “regular” transactions. This mass of transactions is present for all parameter values provided the search cost s is small enough (Theorem 1, part 5). In our model the presence of these transactions reflects moral hazard. Table 1 in Section 2 illustrates the presence of such transactions in the data.

The third implication, contained in Proposition 1, is that if there is a masspoint of transactions with zero quality then there must be a gap in the quality supplied by the market—an interval $(0, R]$ such that no seller offers quality q in this interval. The theory does not tell us how large this gap is. We tested for the presence of a region with zero density just above zero, using the following methodology. We approximate the empirical c.d.f. by a cubic spline with four knots placed at the 15th, 25th, 50th, and 75th percentiles of the distribution in Figure 2. We place the “extra knot” close to zero (the 15th percentile knot) in order to better approximate these functions near the area of interest which is the zero-th percentile. This amounts to regressing the c.d.f. on $1, q, q^2, q^3, (q - k_1)_+^3, \dots, (q - k_4)_+^3$, where k_i are the knot values. For each of the three distributions in Figure 2 we find that coefficient on the linear term q is not significantly different from zero, which means that we cannot reject the hypothesis that, as the c.d.f. approaches zero, it does so with zero slope. A zero slope in the

²³One might take a different view and attribute the dispersion to *unobserved* variation in the attributes of the transactions (safety, convenience, etc.). Undoubtedly some of this variation is present in our data. However, it would be difficult to explain why variation in these unobserved attributes should give rise to the peculiar shape of the quality distributions which are documented next.

c.d.f. corresponds to zero density of the p.d.f., precisely the hypothesis we wanted to test. For details, see Appendix D.

The fourth implication concerns the shape of the quality distribution. Proposition 1 provides the qualitative features of the density distribution of the quality offered by sellers in equilibrium, under the assumption that $\frac{\alpha(1-\gamma)}{\delta+\alpha(1-\gamma)} \leq \frac{3}{4}$. This density has a masspoint at zero, has no mass in the interval $(0, R)$, and then has a decreasing density in the interval $[R, \bar{q}]$. We would like to compare these features of the theoretical distribution with the empirical distributions of pure quantity for the three drugs. As an example, Figure 1 depicts the distribution of pure quantity of crack cocaine traded for \$20 in Washington, DC in the period 1989-1991. This empirical distribution shows some of the features we expect.

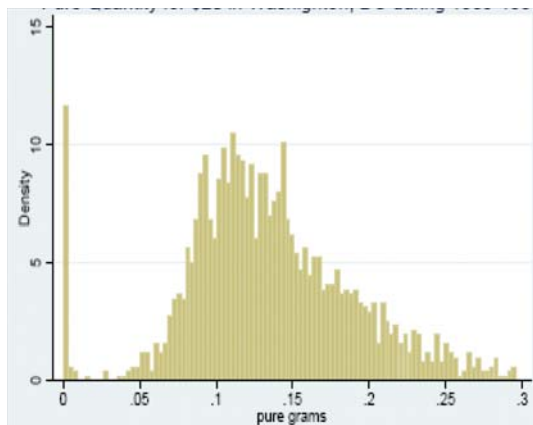


Figure 1: Pure quantity of crack traded for \$20 in Washington DC, 1989-1991.

However, Figure 1 only depicts a narrow slice of the whole market because it only refers to \$20 transactions, it only portrays DC, and it needs to limit the number of years in order to limit the confounding effect of inflation. Most importantly, a picture like Figure 1 would be very difficult to draw for most cities due to the numerosity problem—we just do not have enough observations to draw the equivalent picture for most cities. To deal with these problems, it is necessary to devise a strategy for aggregating many pictures like Figure 1. The first limitation is addressed by studying pure grams *per hundred (of 1983) dollars*, so for example the pure grams bought with \$20 would be multiplied by 5. The second and third limitations require more subtlety. A key problem is that not all years and cities have the same mean quality, due to inflation effects, time trends in purity and wholesale prices, and differences in conditions across cities. Such shifts in the distribution represent a confounding factor for our

purpose, because we are interested in the *shape* of the distribution in a city/year, and not on where it is centered. The (admittedly crude) procedure we use to neutralize the effect of these shifts is to normalize observations by dividing each observation by a city/year average pure quantity. This normalization rescales the horizontal axis of the quality distribution and does not change its qualitative shape. We then aggregate the normalized observations by drugs type, and display the results in Figure 2.²⁴

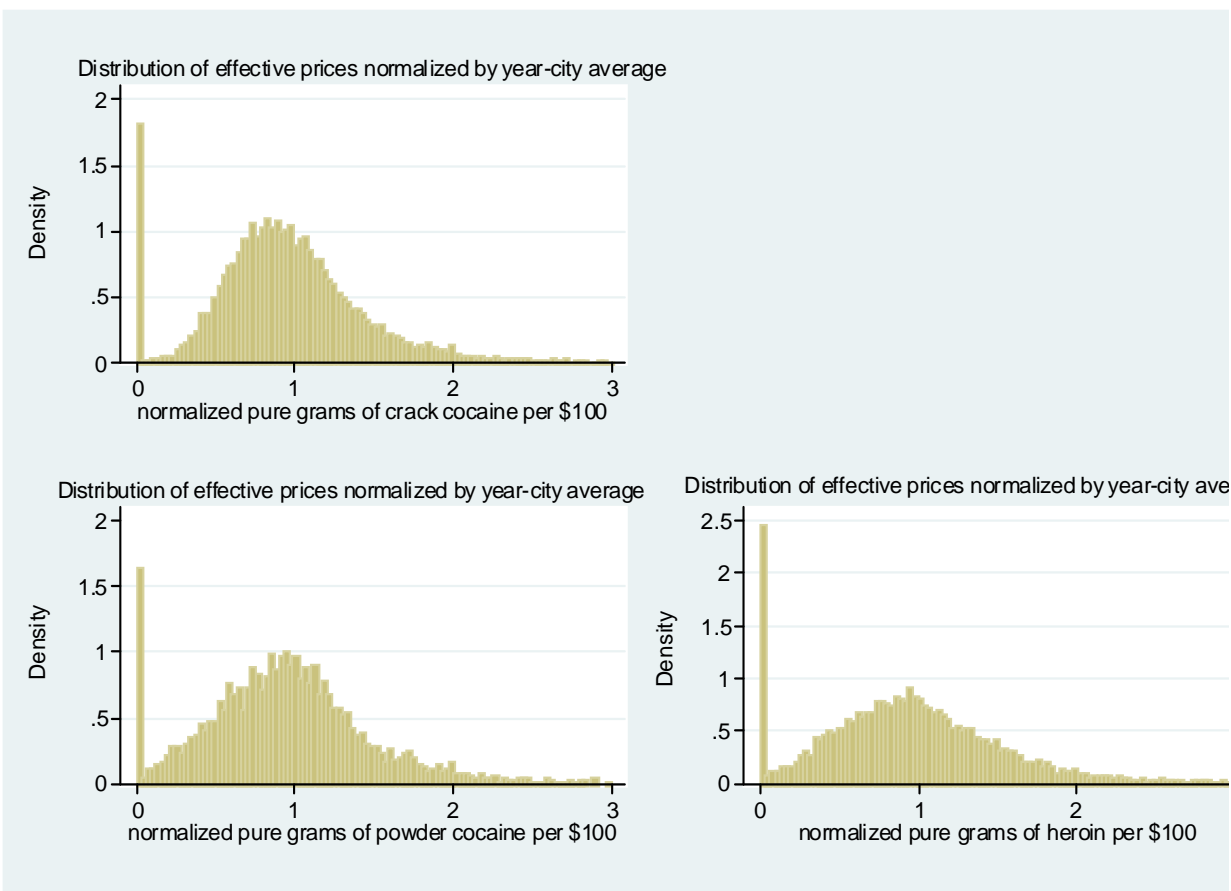


Figure 2: Normalized pure grams per \$100.

Strictly speaking, the model predicts a unimodal distribution within any given market. The empirical distributions in Figure 2 do not show a monotonically decreasing density. We ascribe this non-monotonicity to aggregation issues.²⁵ Because of these issues, it is difficult to find strong support for the hypothesis of decreasing density in Figure 2.

²⁴Figure 5 excludes transactions with value greater than \$100 in 1983 dollars, in order to focus on the retail market. Also, whenever there is only one observation for a city/year, it is dropped. The pictures do not substantially change if all observations are included.

²⁵For example, one might expect some heterogeneity in the search costs across cities. Consider therefore a

The fifth set of implications concerns consumer loyalty. Table 2 in Section 2 illustrates the prevalence of long term relationships and consumer loyalty, which are predicted by our model. That table also suggests that more frequent consumers are more loyal, which would be consistent with a version of our model in which buyers are heterogeneous in their frequency of consumption. Intuitively, this is because conditional on having a regular seller, a buyer with a higher reservation value receives higher quality which makes him more likely to return after a temporary disruption of the match.

The sixth set of implications concerns the correlation between wholesale price and retail affordability. Our model affords sharp comparative statics results concerning the median of F .

Proposition 3 *Suppose fewer than half the sellers rip off their customers. Then a small increase in the wholesale price reduces the median quality offered per amount of money m . Formally, if $F(0) < 1/2$ then $\partial F^{-1}(1/2)/\partial c < 0$.*

Proof. See Appendix A.1. ■

Figure 3 plots the relationship between the (average) wholesale price in a given year and the pure quantity that can be purchased with \$100 at the retail level.²⁶ The figure shows that wholesale (mean) price and retail median (and average) quality are negatively correlated, consistent with the theoretical prediction of Proposition 3. The model’s prediction of a negative correlation between wholesale price and retail quality (median or average) is not especially remarkable: many other models would presumably yield the same correlation. Nevertheless, it is a useful “sanity check” for our model.

standard symmetric, single-peaked distribution of search costs centered around some mode \bar{s} . Ceteris paribus, this heterogeneity will result in c.d.f.’s of quality which differ across cities only in the R ’s and in the size of their masspoints at zero. The corresponding p.d.f.’s will all have decreasing densities; their R ’s (lowest positive quality offered) will be different, and will be centered around the “modal” \bar{R} . After aggregating across cities, we get a distribution whose density is not everywhere decreasing, and which will be qualitatively similar to the empirical distributions in Figure 2.

²⁶ “Retail” transactions are those worth less than \$100, and “wholesale” transactions are those worth more than \$1,000, in 1983 dollars.

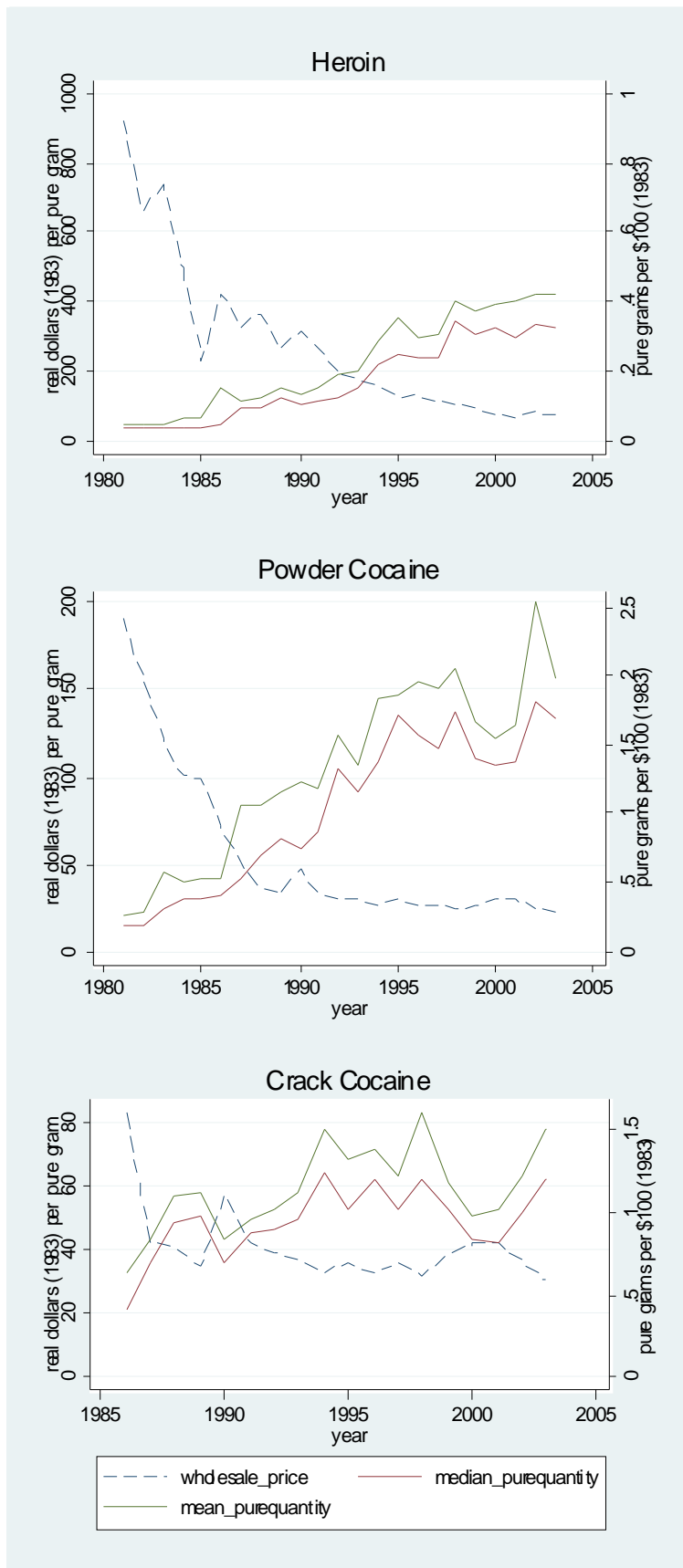


Figure 3: Wholesale price, and retail quality (mean and median).

6 Enforcement and Sentencing

In this section we report a number of comparative statics results which will provide some insight into the effect of various policies aimed at interfering with the retail market. We shall focus on parameter changes related to two types of policies, enforcement and sentencing, and ask what effect these parameter changes have on the affordability of drugs (quality per unit of money m spent).

A foreword on the interpretation of the comparative statics results is in order. Our comparative statics results concern the impact of policy on the terms of trade—how much drugs a given amount of money m can fetch. What impact the terms of trade have on consumption is an open question, the answer to which depends on how the consumer’s “drugs budget” adjusts. In our model the consumer’s “drugs budget” is fixed at m , but in a more general model we could let m , and thus the demand for drugs, adjust to the terms of trade. In such a model we would generally expect that worsening the terms of trade would decrease consumption, particularly for drug users who are not (yet) addicted.²⁷ However, for hard-core addicts with very inelastic demand for drugs, making drugs more expensive may not reduce consumption all that much (or reduce consumption, but not expenditures on drugs). For these hard-core addicts, the kind of policies we consider are more likely to have an impact on consumption (or expenditure) when applied in conjunction with policies designed to reduce the addiction of the hard-core’s consumers (and thus increase their demand elasticity).

6.1 Enforcement

This section studies the comparative statics of our model with respect to a number of parameters. These comparative statics can be thought of as representing the effects of increased enforcement.²⁸

We find that simply deterring some sellers need not *per se* affect the affordability of drugs.

²⁷If m is treated as endogenous, a policy change that, for a fixed m , worsens the terms of trade, might have the possible side effect that consumers might increase the amount spent on drugs from m^* to \tilde{m}^* in order to support their habit. Economic intuition suggests that this effect is unlikely to fully undo the direct effect on the terms of trade—that is, we do not expect quantity consumed to rise as the terms of trade worsen for the buyer. Still, to the extent that m is financed by illegal activities, an increase in m might be an undesirable side-effect of interfering with the market.

²⁸The comparative statics in points 1 and 2 of proposition 4 can also capture changes in sentencing.

The reason is that the remaining sellers may pick up some of the slack. In our model, in fact, this effect completely offsets deterrence and so the direct effect of fewer dealers is nil (Proposition 4 Part 1).

We also find, counterintuitively, that increasing the search cost s results in improved quality of drugs (Proposition 4 Part 2). This is because increasing the search costs makes the buyer less likely to search, and so sellers become more willing to “invest” in a long-run relationship with the buyer instead of settling for the quick rip-off.

A “collateral” effect of increased enforcement is that, as police activity increases, buyer-seller relationships might be temporarily interrupted more frequently. In our model, this effect is captured by an increased temporary separation rates of matches (γ in our model). Proposition 4 Part 3 indicates that increasing the temporary separation rate need not worsen the quality distribution.

Proposition 4 (*Impact of enforcement on terms of trade*)

1. (*Through seller deterrence*) Reducing the number of sellers (or the number of buyers) affects neither the equilibrium quality distribution nor consumer behavior.
2. (*Through reduced consumer search*) As the consumer’s search cost s increases, the average quality of drugs offered by sellers (affordability) for a given m increases and the median does not decrease.
3. (*Through increased temporary break-ups*) As the temporary disruption rate γ increases, the average and median quality of drugs offered by sellers for given m (affordability) may increase.

Proof. See Appendix A.1. ■

Proposition 4 paints a complex picture of the effects of enforcement on equilibrium quality. Increasing enforcement on sellers achieves deterrence, but deterrence has no effect on the affordability of drugs. Increasing s , the buyer’s search cost, improves the quality distribution. We will return to this point when we discuss Proposition 7. Increasing γ , the temporary separation rate is also not necessarily advisable, for at least two reasons. First, Proposition

4 does not give a monotonic prediction, so we may increase γ in a region where doing so actually increases the quality traded. Indeed, such a result could actually help explain why we have seen the average price per pure gram fall during a period of increased enforcement and enforcement budgets. Second, to the extent that increasing the temporary separation rate is achieved through increased enforcement, the direct effect may well be to increase the buyer's welfare, and thus consumer entry in the market. All in all, enforcement affects the quality distribution in complex ways.

6.2 Sentencing of sellers

We now study the effect of sentencing policies for dealers. Sentencing policy in the US takes into account the quantity of drugs that the dealer sells, but not its purity. We now present two sentencing schemes where the sentence does not depend on purity, and a third where it does. We make the simplifying assumption that the quantity is fixed and the same for all trades.

The first scheme is one where a dealer is convicted based on evidence of one trade only (the undercover bust, for example). Then, assuming all trades have the same quantity, all dealers who are caught are put in jail for a period

$$J. \tag{P1}$$

Alternatively, if a dealer is convicted based on the size of his business (perhaps because the police has obtained such evidence via a search), then dealers who trade more go to jail for longer. Let us assume a multiplicative sentence structure where the time spent in jail by a seller of quality q who is caught is given by

$$J \cdot t(q, J), \tag{P2}$$

We denote by $t(q, J)$ the mass of trades made by a seller with quality q in an equilibrium with sentence parameter J . The expression $t(q, J)$ is, of course, determined as part of the equilibrium; the dependence on J arises because J affects the quality distribution offered in equilibrium and therefore the size of a dealer's business who sells quality q .

Finally, we consider an alternative penalty scheme which is not part of the current sentencing guidelines. Under this scheme, the seller gets a mitigation on his sentence if he sells “sub-par” purity relative to the best quality ever traded in equilibrium.²⁹ Formally, we study the following penalty scheme:

$$[J - j(\bar{q} - q)] \cdot t(q, J, j), \tag{P3}$$

The factor j is a parameter representing the intensity of the mitigation. The expression $t(q, J, j)$ represents the mass of trades made by a seller with quality q in an equilibrium with sentence length J and discount j ; it is determined as part of the equilibrium. The number \bar{q} is fixed, for convenience at the upper bound of the quality distribution prevailing when $j = 0$.

The next proposition explores the effects of varying these parameters on the terms of trade and also on the prison population. To this end, we need to introduce into the model the possibility of going to jail. We do this in the simplest possible way, by assuming that a seller is jailed when he meets a first-time customer who is in fact an undercover officer.³⁰ Given a mass 1 of undercover officers who meet with sellers at a constant Poisson rate ζ , the arrest rate for an individual seller is equal to ζ/σ . This outflow of jailed sellers needs to be matched by a corresponding inflow in a stationary equilibrium. It could be new sellers entering, or old sellers coming out of jail—it does not matter for our results.

Proposition 5 (*Effect of sentencing policies on terms of trade and prison population*)

1. *If, as in P1, penalties for dealing are independent of the size of the dealer’s business, then increasing sentences J has no effect on quality per amount of money m (affordability) and consumer behavior, but it increases the jail population.*
2. *If, as in P2, penalties for dealing are increasing in the size of the dealer’s business, then increasing sentences J may help decrease quality per amount of money m (it has*

²⁹In practice, introducing discounts for low purity might drive sellers to work on a margin not studied in this paper: they might decrease purity but keep the overall quality constant per unit of money m by increasing the total (raw) quantity. This practice increases the seller’s exposure, however, because penalties are also a function of quantity (weight) traded, so at the margin we would also expect to see the quality-reduction effect we study in this paper.

³⁰These undercover officers exist: they are the ones who collect much of the data in STRIDE.

the same effect on affordability as increasing c , see Proposition 3), but it increases the jail population.

3. *If, as in P3, penalties for dealing are increasing in the size of the dealer's business, then slightly reducing the sentence of those sellers who sell more diluted drugs (a) has the same effect on affordability as increasing the wholesale price c (see Proposition 3); and (b) it also decreases the jail population.*

Proof. See Appendix A.1. ■

Summing up, intervening through conventional policies such as stiffer sentencing may have little effect on the quality offered in equilibrium, because the buyer's matching rate is unaffected by the magnitude of the mass of sellers. Thus deterring some sellers through harsher sentences has no effect on the quality of drugs traded. In addition, stiffer sentencing has the direct effect of increasing the prison population. An unconventional policy intervention, reducing penalties for dealers who sell diluted drugs, does well on both dimensions. The intuition is straightforward: in a market where the affordability of high quality drugs is determined by the sellers' incentives to dilute, introducing sentencing discounts for diluting will induce sellers to dilute more. Within our model, this unconventional policy has exactly the same effect as an increase in the wholesale price of drugs—a major objective of drugs policy which is pursued through expensive eradication programs in foreign countries, interdiction at entry, etc.—but it is implementable at no significant cost. Moreover, the policy has the added benefit of reducing the prison population.

7 Endogenous demand

In this section we extend the basic model to incorporate an endogenous demand for drugs. Other extensions are relegated to Appendix E.

We consider the case where the steady state number of buyer who participate in the market is endogenously determined through entry and exit. We assume that buyers enter the market unmatched, at a rate that is positively related to the value from being in the market. This means that if quality is high, then a lot of buyers will enter. We show that the mass of buyers is uniquely determined in equilibrium; moreover, for the special case $r = 0$ we are

able analytically to derive comparative statics results of the effect of the parameters on the size of demand.

New buyers enter at rate $\mu(\bar{V})$, where \bar{V} is the value of a newly entered buyer, and $\mu(\cdot)$ is a strictly increasing function. New buyers enter the market unmatched. A buyer exits the market at rate ψ at which point he receives 0. Therefore, if there are β buyers in the market, the flow out of the market is given by $\beta\psi$. We will focus on equilibria where the number of buyers is constant over time. In other words, $\psi\beta^* = \mu(\bar{V})$ is the equilibrium condition that determines the steady state number of buyers, β^* . As in the rest of the paper, we focus on equilibria with a masspoint of sellers offering zero quality.

Proposition 6 *There exists a unique equilibrium with endogenous demand. Given the equilibrium level of demand β^* , the quality distribution has the following properties. First, if q is offered by a seller then either $q = 0$ or $q \in [R, \bar{q}]$, where*

$$\bar{q} = \frac{m}{c} \cdot \frac{\alpha (1 - \gamma)}{\delta + \psi + \alpha (1 - \gamma)}.$$

Moreover, the equilibrium quality distribution is given by

$$F(q) = 1 + \frac{\delta + \psi}{\alpha \gamma} - \frac{1}{\alpha \gamma} \sqrt{\alpha (\delta + \psi) (1 - \gamma)} \sqrt{\frac{(m/c) - q}{q}}$$

on the positive part of its support.

Proof. See Appendix F. ■

The next proposition provides analytical results about the effect of parameter changes on entry. For convenience, it restricts attention to the case $r = 0$, which implies that buyers effectively discount the future at rate ψ .

Proposition 7 *Suppose $r = 0$. Then the equilibrium number of buyers β^* has the following*

properties:

$$\begin{aligned} \frac{d\beta^*}{ds} &< 0 \\ \frac{d\beta^*}{d\delta} &= \frac{d\beta^*}{d\psi} < 0 \\ \frac{d\beta^*}{dc} &< 0 \\ \frac{d\beta^*}{d\alpha} &> 0. \end{aligned}$$

Proof. See Appendix F. ■

Most of the results contained in this proposition are as expected, except for the first. It shows that increasing the buyer's search cost decreases entry. This is interesting when contrasted with Proposition 4 part 2, which shows that increasing search costs increases equilibrium quality. By contrasting these two results we learn that increasing search costs has two opposite effects on buyer welfare: a direct negative effect due to the additional cost of search, and an indirect (general equilibrium) positive effect which comes through the equilibrium quality distribution. Proposition 7 shows that the direct effect is more powerful than the general equilibrium effect, and so increasing the search costs reduces the entry of buyers. An implication of this observation is that, at least in our model, equilibrium quality is not a sufficient statistic for buyer entry: indeed, when search costs shift the two measures move in opposite directions.

In this model, the total amount of drugs consumed is proportional, by construction, to the number of users and to the amount spent on drugs. From a policy perspective, reducing these indicators seems desirable. Proposition 7 reveals the effect of enforcement activities such as increasing the buyers' search cost, or increasing the separation rate between buyers and sellers. As regards the effect of the sentencing schemes analyzed in Section 6.2, the main thrust of the analysis continues to hold in a model with endogenous demand. That is, schemes for sentencing sellers that induce a deterioration of equilibrium quality would reduce the value \bar{V} of a newly entered buyer and thus the equilibrium number of buyers would decrease. In particular, slightly reducing the sentence of those sellers who sell more diluted drugs, would decrease seller demand and also the jail population.³¹

³¹To see this, observe that slightly reducing the sentence of those sellers who sell more diluted drugs has the

8 Conclusions

Over the last 25 years the “war on drugs” has channelled enormous amounts of resources into interfering with the drugs market, yet our view of market structure has largely remained grounded in the Walrasian model. There has not been, to date, an effort to muster a set of empirical regularities in support of an alternative model. The contribution of the paper is to present some new stylized facts about the drugs market and to develop a model that is consistent with them and is informative for policy design. The model is non-Walrasian in that it combines asymmetric information with search theory. A prominent role is played by the ability of sellers to “cut” the drugs without being immediately caught by the customers. This moral hazard puts the retail drugs market at risk of collapse from “overcutting.” The countervailing force that supports trade in our model is the presence of repeated interactions. Despite being stylized, our model matches a number of empirical facts about the drugs market, such as the prevalence of “ripoff” transactions, certain features of the shape of the quality distribution, and the patterns of consumer loyalty.

In the model, a number of conventional enforcement policies can produce counterintuitive outcomes due to general equilibrium effects. If nothing else, these theoretical results underline the importance of understanding the market structure before intervening with policy.

The most intriguing (though speculative) contribution of the model is suggesting unconventional policy interventions. Within our model, a policy of reducing the sentences of sellers who “cheat” and sell low-purity drugs has the same effect as increasing the wholesale price of drugs—a key objective of the war on drugs, and one that is pursued at great cost. In addition, the direct effect of this policy is to reduce incarceration rates relative to current levels. The analysis also suggests that enforcement at the retail level may produce both higher quality of drugs traded and, at the same time, fewer consumers. Enforcement at the wholesale level produces different results.

We do not model many interesting aspects of the illicit drugs markets. First, we abstract from the violence associated with the drugs trade, because we do not have quantitative

same effect on the equilibrium quality distribution as increasing the wholesale price c (refer to Proposition 5 part 3); moreover, we know from Proposition 7 that $\frac{d\beta^*}{dc} < 0$, and this effect comes solely through the equilibrium quality distribution. Thus seller demand decreases. The jail population decreases because not only buyers are being sent to jail for less time, but moreover there are fewer buyers.

evidence of exactly which traits of the market are more or less conducive to violence. We also do not model the possible substitutability across different drugs: *if we only interfere with drug market A*, some of the users who stop using drug A might take up type B. We do not look at the industrial organization of the wholesale sector, nor at possible segmentation among retail markets, nor at non-random consumer search processes. As for consumers, we ignore preference heterogeneity and have only a rudimentary model of time-variation in their preferences; we also assume perfectly rational and optimizing consumers. Finally, we ignore the possibility that there may be switching costs which increase (or perhaps decrease) progressively throughout a buyer-seller relationship. We acknowledge these limitations and hope, if the simple framework presented here is found useful, that it will later be enriched with other realistic features.

We believe the analysis can be relevant in other search markets with moral hazard.

References

- [1] Arkes, J., R.L. Pacula, S. Paddock, J. Caulkins, and P. Reuter. (2004). “Technical Report for the Price and Purity of Illicit Drugs Through 2003.” *Office of National Drug Control Policy*. November.
- [2] Arkes, Jeremy, Susan M. Paddock, Jonathan P. Caulkins, and Peter Reuter (2008). “Why the DEA STRIDE Data are Still Useful for Understanding Drug Markets.” NBER Working Paper 14224, August 2008.
- [3] Becker GS and KM Murphy. 1988. “A Theory of Rational Addiction” *Journal of Political Economy* 96: 675-700.
- [4] Becker, GS; Murphy KM and Grossman M. (2006). “The Market for Illegal Goods: The Case of Drugs” *Journal of Political Economy* 114 (1), pp. 38-60.
- [5] Beggs, Alan and Paul Klemperer. “Multi-Period Competition with Switching Costs” *Econometrica*, Vol. 60, No. 3 (May, 1992), pp. 651-666.
- [6] Berensten, A., and G. Rocheteau (2004). “Money and information”. *Review of Economic Studies* 71 (October), pp. 915–44.
- [7] Burdett, Kenneth and Melvyn G. Coles “Equilibrium Wage-Tenure Contracts” *Econometrica* (2003), 71 (5), 1377-1404.
- [8] Burdett, Kenneth, and Dale T. Mortensen (1998). “Wage Differentials, Employer Size, and Unemployment.” *International Economic Review*, pp. 257-273.
- [9] Burdett, Kenneth, Ricardo Lagos, and Randall Wright (2003). “Crime, Inequality, and Unemployment.” *American Economic Review* 93(5), pp. 1764-1777.
- [10] Bushway, Shawn and Peter Reuter (2008) “Economists’ Contribution to the Study of Crime and the Criminal Justice System”. forthcoming, *Crime and Justice: A Review of Research*.
- [11] Caulkins, J., R.L. Pacula, J. Arkes, P. Reuter, S. Paddock, M. Iguchi, and J. Riley. (2004). “The Price and Purity of Illicit Drugs: 1981 Through the Second Quarter of 2003.” *Office of National Drug Control Policy*. November.

- [12] Caulkins, Jonathan P., Peter Reuter, and Lowell J. Taylor (2006) "Can Supply Restrictions Lower Price? Violence, Drug Dealing and Positional Advantage," *Contributions to Economic Analysis & Policy*, Vol 5(1), 2006.
- [13] Caulkins, Jonathan P., Feichtinger, Gustav; Haunschmied, Josef; and Gernot Tragler. (2006). "Quality cycles and the strategic manipulation of value." *Operations Research* 54(4), pp. 666-677.
- [14] Coles, Melvyn (2001). "Equilibrium wage dispersion, firm size, and growth". *Review of Economic Dynamics* 4(1), pp. 159-187 January 2001.
- [15] Cullum, Paul, and Christopher A. Pissarides (2004) "The Demand for Tobacco Products in the UK." Government Economic Service Working Paper No 150, December 2004.
- [16] Engelhardt, Bryan, Guillaume Rocheteau, and Peter Rupert (2008). "Crime and the labor market: A search model with optimal contracts." Forthcoming, *Journal of Public Economics*.
- [17] Joseph Farrell and Carl Shapiro (1989) "Optimal Contracts with Lock-In" *The American Economic Review*, Vol. 79, No. 1 (Mar., 1989), pp. 51-68.
- [18] Galenianos, Manolis, Rosalie Liccardo Pacula, and Nicola Persico (2009) "A Search-Theoretic Model of the Retail Market for Illicit Drugs." National Bureau of Economic Research Working Paper Series no. 14980.
- [19] Grossman, M. and F.J Chaloupka 1998. The demand for cocaine by young adults: A rational addiction approach. *Journal of Health Economics* 17(4), pp. 427-474.
- [20] Hamid, Ansley (1992). "The Developmental Cycle of a Drugs Epidemic: The Cocaine Smoking Epidemic of 1981-1991." *Journal of Psychoactive Drugs* 24(4), Oct-Dec 1992, pp. 337-348.
- [21] Hoffer, Lee D. (2005) *Junkie Business: The Evolution and Operation of a Heroin Dealing Network* Wadsworth Publishing; 1st edition (December 13, 2005).
- [22] Horowitz, Joel L. (2001). "Should the DEA's Stride Data Be Used for Economic Analyses of Markets for Illegal Drugs?" *Journal of the American Statistical Association*, Vol. 96.

- [23] Huang, Chien-Chieh, Derek Laing, and Ping Wang (2004). "Crime and Poverty: a Search-Theoretic Approach." *International Economic Review*, 45(3), August 2004 , pp. 909-938(30).
- [24] Klemperer, Paul (1995). "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade." *Review of Economic Studies*, 62, 515-539.
- [25] Klemperer, Paul "Markets with Consumer Switching Costs" *The Quarterly Journal of Economics*, Vol. 102, No. 2 (May, 1987), pp. 375-394.
- [26] Mortensen, Dale T. and Tara Vishwanath "Personal contacts and earnings" *Labour Economics* (1994), 1, 187-201.
- [27] Reuter P and J Caulkins (2004) "Illegal Lemons: Price Dispersion in the Cocaine and Heroin Markets" *UN Bulletin on Narcotics*.
- [28] Reuter P. and MA Kleiman (1986). "Risks and prices: An economic analysis of drug enforcement." In M. Tonry & N. Morris (Eds.) *Crime and justice: An annual review of research*, Vol 7, (128-179). Chicago: University of Chicago Press.
- [29] Reuter, Peter, Robert MacCoun, and Patrick Murphy. (1990). "Money from Crime: A study of the economics of drug dealing in Washington, DC." Santa Monica, CA: RAND
- [30] Schelling T 1984. "Self-command in practice, in policy, and in a theory of rational choice." *American Economic Review* 74. pp. 1-11.
- [31] Stigler GJ and GS Becker. 1977. "De Gustibus Non Est Disputandum" *American Economic Review* 67, pp. 76-90.
- [32] Trejos, Alberto (1999). "Search, Bargaining, Money, and Prices Under Private Information." *International Economic Review* 40(3), pp. 679-696
- [33] Wendel, Travis, and Ric Curtis (2000). "The Heraldry of Heroin: 'Dope Stamps' and the Dynamics of Drug Markets in New York City." *Journal of Drugs Issues* 30(2), pp. 225-260.

- [34] Williamson, Steve and Randall Wright (1994). "Barter and Monetary Exchange Under Private Information." *The American Economic Review*, Vol. 84, No. 1 (Mar., 1994), pp. 104-123.

Appendices

A Proofs

A.1 Proof from Sections 3 through 6

Proof of Lemma 1

Proof. We derive equation (3) by using $V(R) = \bar{V}$. It is possible that equation (3) yields a negative reservation value; since quality is nonnegative, in that case we will have $R = 0$.

Using the reservation-value property that we derived, the asset pricing equations can be rewritten as follows:

$$r\bar{V} = \alpha \left[-s + \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \int_R^{\bar{q}} (V(\tilde{q}) - \bar{V}) dF(\tilde{q}) - m \right] \quad (7)$$

$$rV(q) = \alpha \left[(1 - \gamma)q + \gamma(-s + \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \int_q^{\bar{q}} (V(\tilde{q}) - V(q))dF(\tilde{q})) - m \right] + \delta (\bar{V} - V(q)). \quad (8)$$

Recalling that $V(R) = \bar{V}$, equate the two expression above to get (after a bit of algebra)

$$R + s = \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \int_R^{\bar{q}} (V(\tilde{q}) - \bar{V}) dF(\tilde{q}). \quad (9)$$

The next step is to integrate by parts the second integral on the right hand side. We start by calculating $V'(q)$. Differentiate equation (8) with respect to q to get

$$V'(q) = \frac{\alpha (1 - \gamma)}{r + \delta + \alpha \gamma (1 - F(q))}.$$

Now, let us integrate by parts.

$$\int_R^{\bar{q}} (V(\tilde{q}) - \bar{V}) dF(\tilde{q}) = \int_R^{\bar{q}} (1 - F(\tilde{q})) V'(\tilde{q}) d\tilde{q},$$

Substituting the results of this integration into equation (9) yields equation (3). ■

Proof of Proposition 1

Proof. (i) Let n denote the proportion of buyers who are unmatched in steady state. In

steady state, the flow of buyers from the unmatched to the matched state must equal the flow out of the matched state into the unmatched state. An unmatched buyer becomes matched after sampling a seller who offers above-reservation quality which occurs at rate $\alpha (1 - F(R))$. A matched buyer exits the matched state when his match is exogenously destroyed which occurs at rate δ . As a result, in steady state the following holds:

$$\beta n \alpha (1 - F(R)) = \beta (1 - n) \delta \Rightarrow n = \frac{\delta}{\delta + \alpha (1 - F(R))}.$$

The flow of new customers to all sellers consists of unmatched buyers and of matched buyers whose regular supplier is unavailable, that is,

$$\alpha\beta n + \alpha\beta (1 - n) \gamma = \alpha \beta \frac{\delta + \alpha \gamma (1 - F(R))}{\delta + \alpha (1 - F(R))}.$$

Since new customers draw sellers at random, the flow of new customers is equally apportioned among all sellers, and so the per-seller flow of new customers t_N is given by

$$t_N = \frac{\alpha\beta}{\sigma} \frac{\delta + \alpha \gamma (1 - F(R))}{\delta + \alpha (1 - F(R))}$$

This flow of trades is the only business done by sellers who offer $q < R$, and so we have obtained expression (6).

We now solve for $t_L(q)$, the flow of sales from loyal customers. Let $l(q)$ denote the number of loyal customers of a seller offering q . The flow of trades that these buyers generate for a seller offering quality $q \geq R$ is given by

$$t_L(q) = \alpha (1 - \gamma) l(q). \tag{10}$$

We now solve for $l(q)$ (alternatively, we could calculate the expected profit per buyer over the lifetime of the buyer-seller relationship, as done in Burdett and Coles (2003) in a labor market context; our present formulation facilitates the characterization of the quality distribution, whose shape we aim to match).

It is immediate that $l(q) = 0$ when $q < R$. To describe $l(q)$ for $q \geq R$ we need to first characterize G , the c.d.f. of qualities that buyers receive in steady state conditional on being matched. G first order stochastically dominates the (unconditional) distribution of offered qualities because a matched buyer moves to higher qualities over time. The mass of matched buyers receiving quality up to q is given by $\beta (1 - n) G(q)$. The flow into this group comes from the $n\beta$ unmatched buyers who drew a quality level that they chose to keep (i.e. above

R) but which is no greater than q . Note that there are also movements within this group (i.e. from some q_1 to q_2 with $R \leq q_1 < q_2 \leq q$) but these do not affect $G(q)$. Buyers flow out of this group either because their match is exogenously destroyed or because their regular seller was unavailable when they wanted to consume and they sampled a quality level higher than q which made them switch. Equating these flows yields

$$\beta n \alpha [F(q) - F(R)] = \beta (1 - n) G(q) [\delta + \alpha \gamma (1 - F(q))],$$

Isolate $G(q)$ and substitute for n to get

$$G(q) = \frac{\delta [F(q) - F(R)]}{(1 - F(R)) [\delta + \alpha \gamma (1 - F(q))]}$$

for $q \geq R$, and $G(q) = 0$ otherwise.

The mass of buyers matched with a seller who offers quality in $[q - \epsilon, q]$ is $\beta (1 - n) \cdot [G(q) - G(q - \epsilon)]$. The mass of sellers who offer quality in $[q - \epsilon, q]$ is $\sigma [F(q) - F(q - \epsilon)]$. Therefore, the number of buyers matched with a given seller offering quality level q is given by

$$l(q) = \lim_{\epsilon \rightarrow 0} \frac{\beta}{\sigma} (1 - n) \frac{G(q) - G(q - \epsilon)}{F(q) - F(q - \epsilon)} = \frac{\beta}{\sigma} (1 - n) \frac{G'(q)}{F'(q)}$$

(we assume here, and later verify, that F is differentiable for $q > R$). Differentiate $G(\cdot)$ and substitute for n and, after some algebra, we get

$$l(q) = \frac{\alpha \beta}{\sigma} \frac{\delta}{[\delta + \alpha \gamma (1 - F(q))]^2} \frac{\delta + \alpha \gamma (1 - F(R))}{\delta + \alpha (1 - F(R))}$$

Substitute this expression into (10) and add t_N to get the business done by a seller who offers $q \geq R$. This expression corresponds to (5). Note that this derivation did not depend on the existence of a mass point at zero.

(ii) A seller who offers $q \in [0, R)$ has no loyal customers. As result $t(q) = t_N$ for all $q \in [0, R)$ and any positive quality is dominated by $q = 0$.

(iii) Suppose that a discrete mass of sellers offers quality $q^* \geq R$. As a result, there is a mass of buyers whose regular supplier offers q^* . A seller who offers $q^* + \epsilon$ can poach customers from the whole mass of suppliers who offer exactly q^* , leading to a discrete increase in the inflow of buyers. Such a seller would thus get discretely more sales in steady state than any seller offering q^* , with only negligible additional cost and hence $\pi(q^* + \epsilon) > \pi(q^*)$. This cannot hold in equilibrium, and so there cannot be a masspoint at or above R . Since no seller would offer $q \in (0, R)$, F has no mass point on the positive part of its support.

(iv) The argument above proves that there will be quality dispersion unless every seller offers 0. Suppose that $F(0) = 1$, so there is a mass point at zero. Then it is easy to see that lemma 1 implies $R = 0$ and the argument in the proof of part (ii) yields a contradiction.

(v) Suppose there is a gap in the support of F between q_1 and q_2 , where $R \leq q_1 < q_2 \leq \bar{q}$. The sellers offering q_1 and q_2 have exactly the same number of loyal customers since they poach from the same set of competitors and hence $t(q_1) = t(q_2)$. Since it is cheaper to offer q_1 we have $\pi(q_1) > \pi(q_2)$, which cannot be part of an equilibrium. Let \underline{q} be the lowest positive quality on offer. Then $t(R) = t(\underline{q})$ which means that in equilibrium $\underline{q} = R$ for the same reason.

(vii) The analytic expression for the distribution F can be recovered as follows. The profits of sellers offering 0 and q are given by

$$\begin{aligned}\pi(0) &= t(0) m \\ \pi(q) &= t(q) (m - c q), \text{ for } q \geq R.\end{aligned}$$

To solve for the distribution of qualities offered in equilibrium, substitute for $t(\cdot)$ and equate $\pi(0)$ and $\pi(q)$. After some algebra, we get the following function (which, for convenience, is defined on the entire \mathbf{R}_+):

$$\mathbb{F}(q) = 1 + \frac{\delta}{\alpha \gamma} - \frac{1}{\alpha \gamma} \sqrt{\alpha \delta (1 - \gamma)} \sqrt{\frac{(m/c) - q}{q}}. \quad (11)$$

On the interval $[R, \bar{q}]$, the c.d.f. of qualities offered in equilibrium coincides with the function $\mathbb{F}(q)$. Outside of that interval, only zero quality is offered. Formally,

$$F(q) = \begin{cases} \mathbb{F}(q) & \text{for } q \in [R, \bar{q}]. \\ \mathbb{F}(R) & \text{for } q \in [0, R]. \end{cases}$$

(vi) The value of the maximal quality \bar{q} can be recovered by setting $F(\bar{q}) = 1$ and solving. We get

$$\bar{q} = \frac{m}{c} \cdot \frac{\alpha (1 - \gamma)}{\delta + \alpha (1 - \gamma)}.$$

(viii) Differentiating expression (11) yields the density of the qualities offered in equilibrium,

$$f(q) = \frac{1}{2} \sqrt{\frac{(1 - \gamma) \delta}{\alpha \gamma^2}} \cdot \sqrt{\frac{(\frac{m}{c})^2}{(\frac{m}{c} - q) q^3}} \text{ for } q \in [R, \bar{q}]. \quad (12)$$

The function $f(q)$ is a strictly decreasing transformation of $(\frac{m}{c} - q)q^3$. The latter function has a unique (local and global) maximum at $q = \frac{3}{4} \frac{m}{c}$. Therefore, $f(q)$ has a unique minimum at $q = \frac{3}{4} \frac{m}{c}$. The support of F , remember, has upper bound \bar{q} . So if $\bar{q} > \frac{3}{4} \frac{m}{c}$, that is, if $\frac{\alpha(1-\gamma)}{\delta+\alpha(1-\gamma)} > \frac{3}{4}$, then f is U-shaped or monotonically increasing; otherwise, f is monotonically decreasing on its support. ■

Proof of Proposition 2

Proof. 1. Expressions (11) and (3) do not depend on β or σ .

2. Observe from equations (5) and (6) that β/σ enters profits multiplicatively, scaling the total number of transactions $t(q)$. Furthermore, σ does not enter the decision problem of the agents anywhere else. Start with an equilibrium characterized by a quality distribution F . Suppose the mass of sellers σ decreases. If, as σ decreases, the remaining sellers keep offering a quality distribution according to F then profits increase by the same proportion for all sellers. But profits remain equal across sellers, and still no seller has a profitable deviation. This means that it is an equilibrium for the remaining sellers to offer quality distribution F and for buyers to leave their strategy unchanged.

3. Immediate from part 2. ■

Proof of Proposition 3

Proof. Let ϕ denote any quantile in the positive part of the support of the quality distribution (the median, for example). Now increase c (or, equivalently, decrease m), and let $\tilde{\phi}$ denote the same quantile in the new equilibrium distribution. Fix $y > F(0)$ and denote the corresponding quantiles before and after the increase in c by $\phi = F^{-1}(y)$ and $\tilde{\phi} = \tilde{F}^{-1}(y)$. We can write

$$\tilde{\phi} = \tilde{F}^{-1}(y) \leq \tilde{\mathbb{F}}^{-1}(y) < \mathbb{F}^{-1}(y) = F^{-1}(y) = \phi,$$

where the weak inequality reflects the definition of $\tilde{F}(q) = \max[\tilde{\mathbb{F}}(q), \tilde{\mathbb{F}}(\tilde{R})]$; the strict inequality comes from $\partial \mathbb{F}(q)/\partial c > 0$ (cf. equation 11); and the second-to-last equality follows from $y > F(0)$. ■

Proof of Proposition 4

Proof. Part 1. This is simply a restatement of Proposition 2.

Part 2. As characterized in expression (11), the function \mathbb{F} does not depend on s . Therefore, in equation (3) s enters the left-hand side only, and then only as an additive constant. So, as s increases to s' the equilibrium R decreases to R' . Moreover, since $\mathbb{F}(q)$ does not depend

on s or R , as s increases the shape of $F(q)$ is unchanged for $q > R$. Thus, an increase in s results in a stochastically dominant shift of the distribution F .

Part 3. We show that average quality is zero when $\gamma = 0$ and when $\gamma = 1$ which means that it moves non-monotonically for intermediate values of γ . When $\gamma = 0$, a buyer never searches after he is matched with a seller which leads to the Diamond paradox: sellers offer the lowest possible quality level. When $\gamma = 1$, a buyer never returns to a seller he has sampled, which destroys all incentives for sellers to offer positive quality. ■

Proof of Proposition 5

Proof. 1. Suppose the penalty is as in (P1). The number of people entering jail in each instant is given by ζ and the amount of time they spend in jail is J , so the steady state prison population is ζJ . Thus increasing J increases the jail population. To see that J has no effect on market quality and consumer behavior, consider the payoff function that a seller maximizes:

$$\pi(q) = (m - c q) t(q) - \frac{\zeta}{\sigma} J.$$

Note that, crucially, J does not affect the quality choice of the seller. Therefore, increasing penalties effectively increases the entry cost to sellers which leads to fewer sellers in the market, each of whom makes higher monetary profits. The quality distribution is unaffected.

2. Suppose now the penalty is calibrated on the number of trades, as in (P2). The aggregate time spent in jail by a cohort arrested at a point in time is

$$\int J \cdot t(q, J) dF(q, J) = J\beta\alpha$$

The mass of seller who goes to jail in each instant is ζ , so the steady state prison population is now $\zeta J \beta \alpha$ which increases in J .

To see the effect on the quality sold again consider the seller's payoff:

$$\begin{aligned} \pi(q) &= (m - c q) t(q, J) - \frac{\zeta}{\sigma} J t(q, J) \\ &= (\tilde{m} - c q) t(q, J), \end{aligned}$$

where we denote $\tilde{m} = m - \frac{\zeta}{\sigma} J$. We know from Proposition 1 part (vi) that the equilibrium F depends solely on the ratio \tilde{m}/c . Increasing J decreases this ratio. Therefore, increasing J shifts F in the same way as an increase in the wholesale cost of drugs c .

3. Suppose the penalty is as in (P3). The aggregate time spent in jail by a cohort sentenced

at any given instant is

$$\begin{aligned}\mathcal{J} &= \int [J - j(\bar{q} - q)] t(q, J, j) dF(q, J, j) \\ &= (J - j\bar{q})\beta\alpha + j \int q t(q, J, j) dF(q, J, j)\end{aligned}$$

The steady state prison population is therefore $\zeta\mathcal{J}$. Now suppose that we start from $j = 0$ and we contemplate a small increase in j . Its effect on the steady state prison population are given by

$$\begin{aligned}\zeta \left. \frac{\partial \mathcal{J}}{\partial j} \right|_{j=0} &= \zeta \left[-\bar{q}\beta\alpha + \int q t(q, J, 0) dF(q, J, 0) \right] \\ &< \zeta \left[-\bar{q}\beta\alpha + \int \bar{q} t(q, J, 0) dF(q, J, 0) \right] \\ &= \zeta\bar{q}[-\beta\alpha + \beta\alpha] = 0\end{aligned}$$

So the prison population shrinks as we push j slightly above zero. As for the effect of increasing j on the quality distribution, observe that with this penalty the seller's payoff takes the form

$$\begin{aligned}\pi(q) &= (m - c q) t(q) - \frac{\zeta}{\sigma} [J - j(\bar{q} - q)] t(q) \\ &= (\hat{m} - \hat{c} q) t(q),\end{aligned}$$

where we denote $\hat{m} = m - \frac{\zeta}{\sigma} (J - j\bar{q})$ and $\hat{c} = c + \frac{\zeta}{\sigma} j$. We know from Proposition 1 part (vi) that the equilibrium F depends solely on the ratio \hat{m}/\hat{c} . Increasing j decreases this ratio provided that $\hat{m} > \hat{q}\hat{c}$, which must hold because profits must be positive in equilibrium. Therefore, increasing j shifts F in the same way as an increase in the wholesale cost of drugs c . ■

A.2 Existence and Uniqueness of Equilibrium

This section proves that an equilibrium exists and it is unique. An equilibrium is identified by a set of actions by buyers and sellers which are best responses to each other. The cutoff R captures buyer behavior, the function F describes seller behavior. An equilibrium is a

pair R^*, F^* . We first characterize the best response for buyers, next we characterize the best response of sellers, and finally we put them together.

A.2.1 Buyers

Denote

$$H(R|F) \equiv -s - RF(0) + \int_R^{\bar{q}} [1 - F(q) + \alpha(1 - \gamma) \frac{1 - F(q)}{r + \delta + \alpha\gamma(1 - F(q))}] dq. \quad (13)$$

The function $H(R|F)$ is akin to a first-order condition which, for given F , identifies the optimal R . The following lemma provides this result.

Lemma 2 *Suppose F^* and R^* are part of an equilibrium. If $R^* > 0$ then $H(R^*|F^*) = 0$. If $R^* = 0$ then $H(R^*|F^*) \leq 0$.*

Proof. Equation (3) in Lemma 1 reads

$$R = -s + \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \alpha(1 - \gamma) \int_R^{\bar{q}} \frac{1 - F(\tilde{q})}{r + \delta + \alpha\gamma(1 - F(\tilde{q}))} d\tilde{q}$$

Integration by parts implies

$$\begin{aligned} \int_0^{\bar{q}} q dF(q) &= \int_0^{\bar{q}} [1 - F(q)] dq \\ &= R[1 - F(R)] + \int_R^{\bar{q}} (1 - F(q)) dq \\ &= R[1 - F(0)] + \int_R^{\bar{q}} (1 - F(q)) dq, \end{aligned}$$

where the last two equalities follows from the fact that, if F and R are part of an equilibrium, $F(q)$ will be constant between R and zero because selling quality in the range $(0, R)$ is dominated by selling quality zero. Then equation (3) reads $H(R|F) = 0$, and Lemma 1 implies the conclusions. ■

A.2.2 Sellers

We take R to be the threshold above which buyers choose not to search, and in this section we take R as exogenous. We characterize the quality distribution F such that all sellers make the same profits.

Using equations (5) and (6), the profits of a seller who offers q are given by

$$\pi(q) = \frac{\alpha\beta\delta + \alpha\gamma(1 - F(R))}{\sigma\delta + \alpha(1 - F(R))} \left[1 + \frac{\alpha\delta(1 - \gamma)}{[\delta + \alpha\gamma(1 - F(q))]^2} \right] (m - cq) \quad \text{for } q \geq R \quad (14)$$

$$\pi(q) = \frac{\alpha\beta\delta + \alpha\gamma(1 - F(R))}{\sigma\delta + \alpha(1 - F(R))} (m - cq) \quad \text{for } q < R \quad (15)$$

The sellers' optimality condition is that $\pi(q) = \bar{\pi}$ for $q \in \text{supp}F$ and $\pi(q) \leq \bar{\pi}$ for $q \notin \text{supp}F$. We start with some basic properties of F as a function of R .

Lemma 3 *If $R = 0$ then $F(0) = 0$. If $R > 0$ then $F(q) = F(0)$ for $q \in [0, R]$.*

Proof. Suppose $R = 0$. Then there cannot be a mass of sellers at zero, because one of them could offer $q = \varepsilon$ with ε arbitrarily small and gain a discrete mass of customers. This deviation would increase his profits relative to the sellers offering zero, which is inconsistent with equilibrium. Suppose now $R > 0$. Equation (15) implies that no seller will offer a positive quality below R , and therefore in $F(q) = F(0)$ for $q \in [0, R]$. ■

We characterize F as a function of R in two steps. First, we determine whether $F(0)$ is zero, one, or some intermediate value. We then characterize the rest of the quality distribution.

Lemma 4 *For any constellation of parameters, there exist cutoff values \underline{R} and \bar{R} such that:*

$$\begin{aligned} R \geq \bar{R} &\quad \Rightarrow F(0) = 1 \\ R \in (\underline{R}, \bar{R}) &\quad \Rightarrow F(0) \in (0, 1) \\ R \leq \underline{R} &\quad \Rightarrow F(0) = 0 \end{aligned}$$

where the cutoffs only depend on parameters:

$$\begin{aligned} \bar{R} &= \frac{m}{c} \frac{\alpha(1 - \gamma)}{\delta + \alpha(1 - \gamma)} \\ \underline{R} &= \frac{m}{c} \frac{\alpha\delta(1 - \gamma)}{\alpha\delta(1 - \gamma) + (\delta + \alpha\gamma)^2} \end{aligned}$$

Proof. Suppose $R > \bar{R}$. If, by contradiction, $F(0) < 1$ then it must be that $\bar{q} \geq R$ and $\pi(\bar{q}) \geq \pi(0)$. Using equations (14) and (15) the latter condition is equivalent to

$$\bar{q} \leq \frac{m}{c} \frac{\alpha(1 - \gamma)}{\delta + \alpha(1 - \gamma)}.$$

But since $R > \bar{R}$ there can be no \bar{q} that satisfies the previous inequality and also $\bar{q} \geq R$. This contradiction establishes that if $R > \bar{R}$ then $F(0) = 1$.

Conversely, suppose $R < \bar{R}$ and that, by contradiction, $F(0) = 1$. Then a seller could offer $q = R$ and, by equation (14), obtain a profit equal to $\frac{\alpha\beta}{\sigma}[1 + \frac{\alpha(1-\gamma)}{\delta}](m - cR)$. If $R < \bar{R}$, this quantity exceeds $\frac{\alpha\beta}{\sigma}m$, which are the profits from selling $q = 0$. This contradicts the hypothesis that $F(R) = 1$. This contradiction establishes that if $R < \bar{R}$ then $F(0) < 1$.

Suppose now that $R > \underline{R}$ and, by contradiction, $F(0) = 0$. Then $F(R) = 0$ and a seller could offer $q = 0$ and, by equation (15), obtain a profit equal to $\frac{\alpha\beta}{\sigma} \frac{\delta + \alpha\gamma}{\delta + \alpha} m$. If $R > \underline{R}$, this quantity exceeds $\frac{\alpha\beta}{\sigma} \frac{\delta + \alpha\gamma}{\delta + \alpha} [1 + \frac{\alpha\delta(1-\gamma)}{[\delta + \alpha\gamma]^2}](m - cR)$, which by equation (14) are the profits from selling $q = R$. This contradicts the hypothesis that $F(R) = 1$. This contradiction establishes that if $R > \underline{R}$ then $F(0) > 0$.

Finally, suppose that $R < \underline{R}$. First, if $R = 0$ then we know immediately from Lemma 3 that $F(0) = 0$. Let us therefore turn to the case $R > 0$. Suppose by contradiction that $F(0) > 0$. Then we have

$$\frac{m}{c} \frac{\alpha\delta(1-\gamma)}{\alpha\delta(1-\gamma) + [\delta + \alpha\gamma(1 - F(0))]^2} > \frac{m}{c} \frac{\alpha\delta(1-\gamma)}{\alpha\delta(1-\gamma) + [\delta + \alpha\gamma]^2} = \underline{R} > R.$$

Using equations (14) and (15), the inequality between the first and last terms is equivalent to $\pi(0) < \pi(R)$. This is inconsistent with an equilibrium in which $F(0) > 0$. This contradiction establishes that if $R < \underline{R}$ then $F(0) = 0$. ■

Now a full characterization of F , still taking R as exogenously given.

Lemma 5 *The quality distribution as a function of R is given by:*

1. $R \geq \bar{R} \quad \Rightarrow F(q) = 1 \quad \forall q \geq 0$
2. $R \in (\underline{R}, \bar{R}) \Rightarrow F(q) = 1 + \frac{\delta}{\alpha\gamma} - \frac{1}{\alpha\gamma} \sqrt{\alpha\delta(1-\gamma)} \sqrt{\frac{(m/c) - q}{q}}, \quad q \in [R, \bar{q}]$
 $F(q) = 1 + \frac{\delta}{\alpha\gamma} - \frac{1}{\alpha\gamma} \sqrt{\alpha\delta(1-\gamma)} \sqrt{\frac{(m/c) - R}{R}}, \quad q < R$
 $\bar{q} = \frac{m}{c} \frac{\alpha(1-\gamma)}{\delta + \alpha(1-\gamma)}$
3. $R \leq \underline{R} \quad \Rightarrow F(q) = 1 + \frac{\delta}{\alpha\gamma} (1 - \sqrt{h(q)}), \quad q \in [R, \bar{q}]$
 $F(q) = 0, \quad q < R$
 $\bar{q} = \left[\frac{m}{c} + \frac{\delta R}{\alpha(1-\gamma)} - \frac{m/c - R}{(1 + \alpha\gamma/\delta)^2} \right] / \left(1 + \frac{\delta}{\alpha(1-\gamma)} \right)$
 $h(q) = \frac{m/c - q}{\frac{\delta(q-R)}{\alpha(1-\gamma)} + \frac{m/c - R}{[1 + \alpha\gamma/\delta]^2}}$

Proof. The first part is obvious. The second part is a restatement of proposition 1. For the third part set $\pi(R) = \pi(q)$ and solve for $F(q)$ using equation (14) and setting $F(0) = 0$. ■

This completes the characterization of the quality distribution $F(\cdot|R)$ for an arbitrary R .

A.2.3 Equilibrium

We have reduced the problem of finding an equilibrium to that of identifying a value of R giving rise to an $F(q|R)$ which, when plugged into $H(R|\cdot)$, achieves a value of zero for H . When such an R is unique then we have a unique equilibrium. If such a value cannot be found and H is negative for all R , then $R = 0$ is an equilibrium as per Lemma 1.

Define

$$\hat{H}(R) \equiv H(R|F(\cdot|R)),$$

which is the function we are seeking the zero of.

Lemma 6 $\hat{H}(R)$ has the following features:

1. The function $\hat{H}(\cdot)$ is continuous.
2. When $R \geq \bar{R}$ we have $\hat{H}(R) = -s < 0$ and $d\hat{H}/dR = 0$.

3. When $R \in (\underline{R}, \overline{R})$ we have $d\hat{H}/dR < 0$.

4. When $R \in [0, \underline{R}]$ we have $d\hat{H}/dR < 0$.

5. The function $\hat{H}(\cdot)$ is decreasing in s .

Proof. Continuity of $\hat{H}(\cdot)$ will follow if we can establish continuity of $F(\cdot|R)$. The area of possible concern is when $R = \underline{R}$; in this neighborhood Lemma 5 provides two apparently different expressions for F , one for $R < \underline{R}$, the other for $R > \underline{R}$. However, it can be shown by direct substitution that, at $R = \underline{R}$ the two different expressions coincide. Hence continuity of $F(\cdot|R)$ is established.

Part 2 follows from the characterization of $F(\cdot|R)$.

For part 3, notice that the integral only depends on R through its minimal positive quality. Therefore as R increases the value of the integral in equation (13) decreases. Furthermore, $RF(0)$ is equal to $RF(R)$ and hence is increasing in R (it is the product of two positive and increasing functions of R). This proves the claim.

For part 4, substitute the exact form for F inside the expression for \hat{H} (noting that $F(0) = 0$):

$$\hat{H}(R) = -s + \frac{\delta}{\alpha\gamma} \int_R^{\bar{q}} \left[\sqrt{h(q)} - 1 + \alpha(1-\gamma) \frac{(\sqrt{h(q)} - 1)}{r + \delta\sqrt{h(q)}} \right] dq.$$

Perform the change of variables: $w = h(q)$ which implies that

$$\hat{H}(R) = -s - \frac{\delta}{\alpha\gamma} \int_{\underline{w}}^{\bar{w}} \left[\sqrt{w} - 1 + \alpha(1-\gamma) \frac{(\sqrt{w} - 1)}{r + \delta\sqrt{w}} \right] \left(\frac{dq}{dw} \right) dw.$$

where $\bar{w} = h(\bar{q}) = 1$ and $\underline{w} = h(R) = (1 + \alpha\gamma/\delta)^2$. Now,

$$q = h^{-1}(w) = \frac{\alpha(1-\gamma)}{\alpha(1-\gamma) + w\delta} \left\{ (m/c) + w \left[\frac{\delta R}{\alpha(1-\gamma)} - \frac{(m/c) - R}{[1 + \alpha\gamma/\delta]^2} \right] \right\}.$$

Differentiate with respect to w and collect terms to get

$$\left(\frac{dq}{dw} \right) = - \frac{m/c - R}{(1 + \frac{\delta w}{\alpha(1-\gamma)})^2} \left[\frac{\delta}{\alpha(1-\gamma)} + \frac{1}{(1 + \alpha\gamma/\delta)^2} \right].$$

Therefore, we have

$$\begin{aligned} & \hat{H}(R) \\ &= -s + \frac{\delta}{\alpha\gamma} \left(\frac{m}{c} - R \right) \left[\frac{\delta}{\alpha(1-\gamma)} + \frac{1}{(1+\alpha\gamma/\delta)^2} \right] \int_{\bar{w}}^{\underline{w}} \left(\sqrt{w} - 1 + \alpha(1-\gamma) \frac{(\sqrt{w}-1)}{r+\delta\sqrt{w}} \right) \frac{1}{\left[1 + \frac{\delta w}{\alpha(1-\gamma)}\right]^2} dw \end{aligned}$$

Note that the integral is independent of R . Furthermore, the integrand is positive: indeed, from the expression for F in Lemma 5 it follows that $\sqrt{h(\underline{q})} > 1$ hence $\sqrt{w} > 1$; this, together with $\underline{w} > \bar{w}$, implies that the integral is a positive number. Finally, the term before the integral decreases in R which shows that the function $\hat{H}(R)$ is decreasing in R .

Part 5 follows directly from the definition of $\hat{H}(\cdot)$, given that $F(\cdot|R)$ is independent of s . ■

The previous lemma shows that $\hat{H}(R)$ is decreasing on $[0, \infty)$ and it assumes negative value after \bar{R} . Based on these properties, a unique equilibrium exists, as detailed in the next theorem.

Theorem 1 *For each parameter constellation, the equilibrium pair R^*, F^* exists, it is unique, and furthermore:*

1. *If $\hat{H}(0) \leq 0$, then $R^* = 0$ and $F^*(0) = 0$.*
2. *If $\hat{H}(0) > 0$ and $\hat{H}(\underline{R}) \leq 0$, then $R^* > 0$ and $F^*(0) = 0$.*
3. *If $\hat{H}(\underline{R}) > 0$, then $R^* > 0$ and $F^*(0) > 0$.*
4. *The equilibrium value R^* is decreasing in the search cost s .*
5. *The set of values of s such that an equilibrium with a mass point at zero exists, is a non-empty interval.*

Proof. $\hat{H}(R)$ is strictly decreasing in $[0, \bar{R}]$ and flat thereafter at $\hat{H}(R) = -s < 0$. The monotonicity of $\hat{H}(R)$ leads to uniqueness: it can cross zero at most once. The characterization in parts 1-3 follows immediately. Part 4 follows directly from part 5 of Lemma 6. Part 5 (set of values of s is an interval) follows from combining part 4 with Lemma 4. We now establish the non-emptiness of this interval. Define

$$\begin{aligned} \bar{H}(R) &= s + \hat{H}(R) \\ \bar{s} &= \bar{H}(0) \\ \underline{s} &= \bar{H}(\underline{R}), \end{aligned}$$

We know that $\bar{H}(R)$ and \underline{R} are independent of s (proof of Lemma 6, part 5 and definition in Lemma 4). Recall that depending on the root of \hat{R} there are three possible types of equilibria: (E1) $F(0) = R = 0$; (E2) $F(0) = 0, R > 0$; (E3) $F(0) > 0, R > 0$. Rewriting the statement in parts 1-3 above we get

$$\begin{aligned}\hat{H}(0) \leq 0 &\Leftrightarrow s \geq \bar{s} \Rightarrow E1 \\ \hat{H}(0) > 0 \text{ and } \hat{H}(\underline{R}) \leq 0 &\Leftrightarrow s \in [\underline{s}, \bar{s}) \Rightarrow E2 \\ \hat{H}(\underline{R}) > 0 &\Leftrightarrow s < \underline{s} \Rightarrow E3\end{aligned}$$

where the cutoffs \underline{s} and \bar{s} depend on all parameters other than s . Therefore, given any constellation of parameters, we can find s small enough such that there is a mass point at zero. ■

B The Data

We use information available in the 1981-2003 STRIDE which has a total of approximately 780,000 observations for a number of different drugs and acquisition methods. We have approximately 115,000 observations for heroin and 330,000 for cocaine (crack and powder). We keep the observations acquired through purchases and clean the data of missing values, observations whose weight is lower than 0.1 gram because for those observations the purity is unreliable, and other unreliable observations, as suggested in Arkes et al (2004). We are left with 29,181 observations for heroin, 47,743 for crack cocaine and 46,050 for powder cocaine which we use. The STRIDE data is essentially an administrative dataset collected by police agencies, not a random sampling of drug prices. The reliability of the STRIDE data set has been called into question by Horowitz (2001), who remarked that depending on which agency collected the data (DEA or other law enforcement agency), the time series of drug prices in Washington, D.C. look somewhat different. However, Arkes et al. (2008) show that the inconsistencies identified by Horowitz (2001) largely disappear simply by controlling for the size of the transaction (above or below 5 grams) when combined with other data cleaning issues raised by Horowitz (2001). Mindful of this finding, we are careful to restrict our analysis to the relatively narrow sample of transactions whose value is below 100 constant 1983 dollars. Also, Arkes et al. (2008) show that the price series for different drugs obtained from STRIDE predict, in a Granger sense, the number of drug-related admissions to emergency rooms (DAWN data set). Overall, we feel that Arkes et al. (2008) make a compelling case for the usefulness of the STRIDE dataset when used carefully, i.e., without aggregating across transactions of vastly different sizes.

The ADAM data set is collected quarterly from interviews with persons arrested or booked on local and state charges within the past 48 hours in various ADAM metropolitan areas in the United States. The number of these areas changes from years to year based on the availability of the data.³² Individuals involved in non-drug and drug-related crimes are interviewed with the goal of obtaining information about the use, importance and role of drugs and alcohol among those committing crimes. Since 2000 a probability based sampling design is applied during sampling male arrestees. For the female arrestee a purposive sampling is applied. The arrestees participated in the survey voluntarily under full confidentiality.³³ In addition to interviewing arrestees, urine samples are requested and analyzed for validation of self-reported drug use. Since 2000, a drug market procurement module has been included as part of the quarterly survey and collects information on the arrestee's most recent drugs purchase

³²From 2000 to 2003, it has been 35, 33, 36 and 39, respectively.

³³Dave (2007) notes that only about 10% of the arrestees reject the interview request.

for all arrestees who report having used drugs in the previous 30 days. Information collected includes number of times drugs were purchased in past 30 days, number of drug dealers they transacted with, whether they last purchased from their regular dealer, difficulties experienced in locating a dealer or buying the drug, and the price paid for the specific quantity purchased. The data collection was interrupted in 2003. This paper uses data from 2000 to 2003.

C Quality Dispersion

Below we explore quality dispersion in our sample. We restrict attention to trades that are less than \$100 in 1980 dollars. To interpret Table 4, let us start by focusing on the sub-tables titled “UnR,” and the rows with the legend “Total.” The acronym “UnR” refers to the “unrestricted” sample of all trades that are less than \$100 in 1980 dollars. Row “Total” refers to the entire sample, not broken down into quartiles. So row “Total” in sub-table “UnR” tells us how much pure quantity \$100 buys. In our heroin sample, for example, \$100 buys on average 0.3 pure grams of heroin with a very large standard deviation of 0.35, which leads to a coefficient of variation (standard deviation over mean) of 1.19. This is a very large coefficient, supporting our argument that dispersion is very sizable.

Of course, one might expect that any temporal or geographical difference in prices may inflate our measure of dispersion. For this reason we conduct two fixed effects regressions, with the purpose of controlling for such differences. In both regressions we have city fixed-effects, a time dummy, and a city*time interaction term. In the first regression (denoted by FE1), the time variable is the year; in the second (denoted by FE2), it is the quarter that the transaction took place. We then focus on the residuals of these regressions, the statistics of which are reported in Tables FE1 and FE2. Before conducting these fixed effects regressions, however, we tailor our sample to each fixed effect regression, which is necessary to achieve a reasonable numerosity. For the first fixed-effect regression (FE1) we restrict our sample to cities that have more than 400 observations in total, and we drop city*years in which there are less than 5 observations; this gives rise to the sample described in Table R1. For the second fixed-effect regression, we restrict attention to cities that have more than 950 observations and drop city*quarters in which there are less than 5 observations; this sample, which is described in Table R2, is the basis of the fixed-effect regression FE2. We carry out the fixed effect regressions on each of these samples, and we then compute the coefficient of variation, using the standard deviation of the residuals divided by the sample mean without fixed effects. The results are reported in sub-tables FE1 and FE2.

Heroin					Crack					Powder					
UnR	count	mean	sd	cv	UnR	count	mean	sd	cv	UnR	count	mean	sd	cv	
1	3189	0.15	0.15	1.02	1	4149	0.69	0.43	0.63	1	1342	0.53	0.63	1.19	
2	3198	0.24	0.24	0.98	2	3956	0.9	0.56	0.62	2	1341	0.71	0.66	0.93	
3	3156	0.35	0.31	0.89	3	4059	1.02	0.62	0.61	3	1339	1.06	0.92	0.87	
4	3178	0.44	0.52	1.18	4	4038	1.97	2.67	1.35	4	1340	2.57	3.11	1.21	
Total	12721	0.3	0.35	1.19	Total	16202	1.15	1.5	1.3	Total	5362	1.22	1.87	1.54	
R1					R1					R1					
count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv
1	1815	0.16	0.15	0.91	1	2138	0.68	0.28	0.4	1	310	0.42	0.3	0.72	
2	1773	0.26	0.22	0.84	2	2110	0.91	0.36	0.4	2	311	0.52	0.36	0.69	
3	1774	0.37	0.29	0.79	3	2119	1.03	0.56	0.54	3	307	0.46	0.33	0.71	
4	1785	0.48	0.48	1.01	4	2119	1.55	2.34	1.51	4	309	0.78	0.59	0.75	
Total	7147	0.32	0.33	1.05	Total	8486	1.04	1.26	1.21	Total	1237	0.55	0.43	0.8	
FE1					FE1					FE1					
count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv
1	1815	0	0.1	0.63	1	2138	0	0.24	0.35	1	310	0	0.26	0.62	
2	1773	0	0.16	0.62	2	2110	0	0.33	0.36	2	311	0	0.31	0.60	
3	1774	0	0.19	0.51	3	2119	0	0.52	0.50	3	307	0	0.27	0.59	
4	1785	0	0.35	0.73	4	2119	0	2.29	1.48	4	309	0	0.49	0.63	
Total	7147	0	0.22	0.69	Total	8486	0	1.19	1.14	Total	1237	0	0.34	0.62	
R2					R2					R2					
count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv
1	734	0.16	0.14	0.84	1	1914	0.67	0.26	0.38	1	294	0.43	0.3	0.69	
2	744	0.2	0.17	0.84	2	1894	0.9	0.32	0.36	2	292	0.51	0.35	0.69	
3	726	0.3	0.25	0.82	3	1860	1.02	0.54	0.53	3	295	0.46	0.32	0.7	
4	719	0.37	0.47	1.26	4	1888	1.4	2.1	1.5	4	291	0.76	0.51	0.68	
Total	2923	0.26	0.3	1.15	Total	7556	1	1.13	1.14	Total	1172	0.54	0.4	0.74	
FE2					FE2					FE2					
count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv	count	mean	sd	cv
1	734	0	0.09	0.56	1	1914	0	0.25	0.37	1	294	0	0.26	0.60	
2	744	0	0.11	0.55	2	1894	0	0.32	0.36	2	292	0	0.3	0.59	
3	726	0	0.14	0.47	3	1860	0	0.53	0.52	3	295	0	0.27	0.59	
4	719	0	0.35	0.95	4	1888	0	2.1	1.50	4	291	0	0.45	0.59	
Total	2923	0	0.2	0.77	Total	7556	0	1.1	1.10	Total	1172	0	0.33	0.61	

Table 3: Pure grams that \$100 dollars buys, by quartile of the non-pure amount.

Both restricted samples R1 and R2 have very similar “Total” summary statistics to the UnRestricted sample. Turning to the summary statistics of the residuals (sub-tables FE1 and FE2), we see that the “Total” coefficient of variation remains large (often above 1) even after adding city/time/interaction fixed effects. This suggests that most of the price variation occurs within a point in space and time, rather than across different points.

In order to better understand the dispersion within the sample, we also break down transactions by non-pure weight. Rows 1-4 in each sub-table in Table 4 present the summary

statistics for each of the 4 quartiles of the sample, by non-pure weight. The coefficients of variation are reduced but remain large, indicating that a large amount of dispersion persists even after we break down transactions by (non-pure) weight. We also find that bigger transactions are associated with a higher pure-gram-per-dollar amount, which can be viewed as evidence of “quantity discounts.”

D Spline Test

We devised the following test to detect the presence of a region with zero density. We approximate the empirical c.d.f. by a cubic spline with four knots placed at the 15th, 25th, 50th, and 75th percentiles of the distribution in Figure 2. We chose to place an extra knot close to zero (the 15th percentile knot) in order to better approximate these functions near the area of interest which is the zero-th percentile. This amounts to regressing the c.d.f. on $1, q, q^2, q^3, (q - k_1)_+^3, \dots, (q - k_4)_+^3$, where k_i are the knot values. Figure 4 displays the empirical c.d.f.s and the fitted spline regressions.

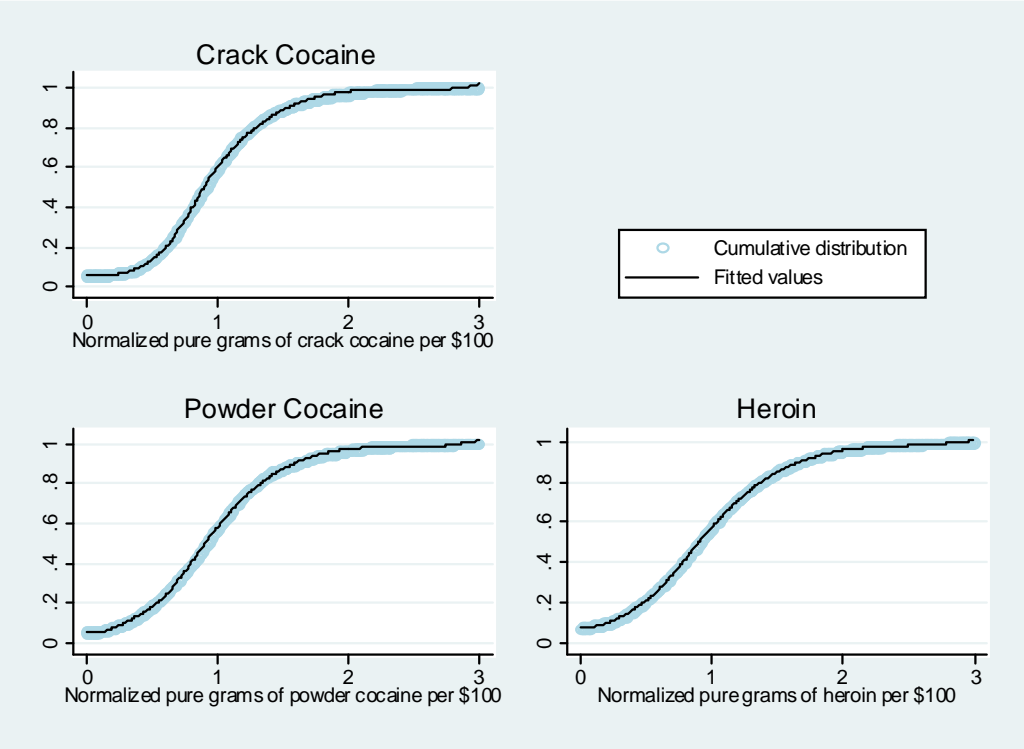


Figure 4: Spline approximation to the empirical c.d.f.’s of the pure quantity distributions.

The coefficient on the linear term q (standard errors in parentheses) equals $-0.036 (0.03)$ for Heroin, $-0.016 (0.04)$ for Crack Cocaine, and $-0.045 (0.5)$ for Powder Cocaine. None of these coefficients is significantly different from zero, which means that we cannot reject the hypothesis that, as the c.d.f. approaches zero, it does so with zero slope. Of course, zero slope in the c.d.f. corresponds to zero density of the p.d.f., precisely the hypothesis we wanted to test.

	Heroin	Crack Cocaine	Powder Cocaine
q	-0.036 (0.03)	-0.016 (0.04)	-0.045 (0.05)
q^2	0.839*** (0.12)	0.033 (0.14)	0.867*** (0.21)
q^3	-0.486** (0.15)	0.626*** (0.14)	-0.522* (0.24)
$(q - k_1)_+^3$	1.421*** (0.27)	0.909 (0.51)	2.575*** (0.54)
$(q - k_2)_+^3$	-2.014*** (0.16)	-5.196*** (0.55)	-3.845*** (0.42)
$(q - k_3)_+^3$	0.934*** (0.05)	4.194*** (0.22)	1.693*** (0.15)
$(q - k_4)_+^3$	0.260*** (0.01)	-0.385*** (0.06)	0.245*** (0.05)
cons	0.075*** (0.00)	0.056*** (0.00)	0.051*** (0.00)

Table 4: Coefficients on spline regression (standard error in parenthesis)

We conducted some robustness tests by introducing an additional knot at the 10th percentile of the quality distribution; the coefficients on q remained not statistically significant for all three drugs. The insignificance persisted when we replaced the 15th percentile knot with a 10th percentile knot.

FOR ONLINE PUBLICATION

E Extensions

E.1 Search Costs Depend on the Number of Dealers

In this section we extend our analysis to the case in which search costs depend on the number of dealers which are present in equilibrium. The idea is that if there are fewer sellers in the market, it may be more difficult for a buyer to physically locate a seller. Due to this feature, the equilibrium quality distribution is no longer independent of the buyer/seller ratio; indeed, the latter affects the former through the search cost. In this section we study how an exogenous change in the propensity of sellers to enter the market affects the equilibrium quality distribution.

We model the effect of the density of dealers by letting the search cost be a decreasing function $s(\cdot)$ of the equilibrium number of sellers σ^* . A possible microfoundation of this decreasing function might be given as follows. Suppose that each time a buyer searches, he incurs a cost s and finds a seller with probability p . If he finds the seller, he transacts; if not, then he searches again. In that case, whenever the seller has the urge to consume his overall expected search cost is given by $sp + 2sp(1 - p) + 3sp(1 - p)^2 + \dots = s/p$. Now suppose $p = p(\frac{\beta}{\sigma^*})$, where we assume that $p(\cdot)$ is a decreasing function of its argument because it is more difficult for a buyer to physically locate a seller when sellers are scarce relative to buyers. Then the buyer's expected search cost is given by $s/[p(\frac{\beta}{\sigma^*})]$, which is decreasing in σ^* .

Proposition 8 *When the search costs are a decreasing function of the number of sellers on the market, an exogenous drop in the sellers' entry cost results in: (a) a greater increase in the mass of sellers compared to our baseline model; and (b) an increase in the fraction of trades which are rip-offs.*

Proof. If the search cost endogenous, then the buyer/seller ratio β/σ^* is no longer independent of the quality distribution. This is because σ^* affects the endogenous search cost $s(\sigma^*)$ which in turn affects the distribution of qualities through the buyer's search threshold R . We know from the proof of Proposition 4 that the equilibrium R is decreasing in s , which implies that the equilibrium R is increasing in σ^* . So an exogenous increase in the propensity of sellers to enter the market, due say to a decrease in the entry cost K , will have two

partial equilibrium effects. First, the direct effect of increasing σ^* to some $\sigma' > \sigma^*$ which will decrease the search cost. Second, due to the decrease in the exogenous search cost, R will increase to some R' . This second effect of course changes the quality distribution.

To put together these two partial equilibrium effects, we need to see how the second effect impacts the first (we need not worry about the direct impact of σ^* on R because there is none). To this end, observe that fixing the number of sellers to σ^* and moving to $R' > R$ increases sellers' profits. This is because there will be more rip-off sellers and therefore the seller selling the highest quality will have a greater influx of new customers. But if the highest quality seller increases his profits, then all sellers do too, because all sellers make the same profits in equilibrium. This shows that raising R increases the dealer's incentives to enter the market.

Putting together these two partial equilibrium effects, we conclude that an exogenous increase in the dealers' propensity to enter has two effects: a new one, which is to worsen the quality distribution by increasing the fraction of rip-off dealers, at the expense of low-but-positive quality dealers; and the old effect, which is to increase the equilibrium number of dealers. But the magnitude of the old effect is now amplified by the changes in the quality distribution, which have the effect of increasing the dealers' propensity to enter. ■

Notice that, in our baseline model, effect (b) is not present because the quality distribution was independent of the buyer/seller ratio. The lesson to be drawn from this analysis is that, when search costs vary with dealer scarcity, an exogenous drop in the dealers' cost to enter the market generates more dealers on the market *and* more rip-offs. More generally, we can say that in this extension, an exogenous variation in supply generates a negative correlation between supply size (number of dealers) and quality. This relationship is not present in the baseline model.

E.2 Determination of the wholesale cost of drugs

Our model has focused on the retail side of the business. We can embed this model into a more general model of demand and supply, by assuming that the wholesale cost of drugs (c in our model) is determined by an increasing supply curve which reflects the global technology for producing drugs. This model allows the marginal cost to be pinned down by the equilibrium condition

$$S(c) = D(c),$$

where $S(c)$ is an exogenous function representing supply of pure drugs. Supply depends on c , the price that the wholesaler receives. We make the standard assumption that the function $S(\cdot)$ is increasing. On the other side we have demand for pure drugs

$$D(c) = \beta^* \cdot \int_0^\infty q \cdot t(q) dF(q).$$

Under some fairly standard assumptions on the supply function ($S(0) = \infty, S(\infty) = 0$) existence of an equilibrium value for c^* is guaranteed. Given c^* , the model plays out as described in the previous sections.

F Proofs from Section 7

Proof of Proposition 6

Proof. To prove the proposition, we first replicate the analysis of Section 3, adding the new feature of buyer entry and exit. We rewrite the buyers' value functions, now taking into account the possibility of exit:

$$\begin{aligned}
 r\bar{V} &= \alpha[-s + \int_0^{\bar{q}} dF(\tilde{q}) + \int_R^{\bar{q}} (V(\tilde{q}) - \bar{V})dF(\tilde{q}) - m] - \psi \bar{V} \\
 rV(q) &= \alpha[(1 - \gamma)q + \gamma (-s + \int_0^{\bar{q}} \tilde{q}dF(\tilde{q}) + \int_q^{\bar{q}} (V(\tilde{q}) - V(q))dF(\tilde{q}) - m] \\
 &\quad + \delta(\bar{V} - V(q)) - \psi V(q).
 \end{aligned} \tag{16}$$

Equating $V(R) = \bar{V}$ yields

$$R = -s + \int_0^{\bar{q}} dF(\tilde{q}) + \int_R^{\bar{q}} (V(\tilde{q}) - \bar{V})dF(\tilde{q})$$

which, when combined with equation (16) leads to:

$$(r + \psi)\bar{V} = \alpha(R - m) \tag{17}$$

We shall use this expression to show that \bar{V} is unique and does not depend on the number of buyers and sellers.

Performing similar calculations to lemma 1, the buyers' reservation quality is given by

$$R = -s + \int_0^{\bar{q}} \tilde{q} dF(\tilde{q}) + \alpha (1 - \gamma) \int_R^{\bar{q}} \frac{1 - F(\tilde{q})}{r + \psi + \delta + \alpha \gamma (1 - F(\tilde{q}))} d\tilde{q} \tag{18}$$

We now compute an expression for the equilibrium F and show that, as in the previous analysis, it is independent of the buyer/seller ratio.

In steady state the flows to and from the matched state are equal and therefore:

$$\beta n \alpha (1 - F(R)) = \beta (1 - n) (\delta + \psi)$$

where the only difference with our earlier expression is that buyers leave their seller at additional rate ψ . We can similarly rewrite the flows of loyal customers into and out of the

group of sellers who offer quality up to q :

$$\beta n \alpha [F(q) - F(R)] = \beta (1 - n) G(q) [\delta + \psi + \alpha \gamma (1 - F(q))].$$

Therefore, the steady state level of transactions of a seller offering quality q is given by

$$t(q) = \frac{\alpha\beta}{\sigma} \frac{\delta + \psi + \alpha \gamma (1 - F(R))}{\delta + \psi + \alpha (1 - F(R))} \left[1 + \frac{\alpha (\delta + \psi) (1 - \gamma)}{[\delta + \psi + \alpha \gamma (1 - F(q))]^2} \right], \quad \text{when } q \geq R,$$

$$t(q) = \frac{\alpha\beta}{\sigma} \frac{\delta + \psi + \alpha \gamma (1 - F(R))}{\delta + \psi + \alpha (1 - F(R))}, \quad \text{when } q < R.$$

Replicating the calculations from the main model, one finds that there is a unique equilibrium and when $F(0) > 0$ the distribution of qualities for $q \in [R, \bar{q}]$ is given by

$$F(q) = 1 + \frac{\delta + \psi}{\alpha\gamma} - \frac{1}{\alpha\gamma} \sqrt{\alpha(\delta + \psi)(1 - \gamma)} \sqrt{\frac{(m/c) - q}{q}}, \quad q \in [R, \bar{q}] \quad (19)$$

$$F(q) = 1 + \frac{\delta + \psi}{\alpha\gamma} - \frac{1}{\alpha\gamma} \sqrt{\alpha(\delta + \psi)(1 - \gamma)} \sqrt{\frac{(m/c) - R}{R}}, \quad q \in [0, R] \quad (20)$$

Note that the quality distribution does not depend on the buyer/seller ratio and therefore, from (18), the buyers' reservation R does not depend on it either. It follows from expression (17) that the value \bar{V} is determined independently of β and σ . Hence $\mu(\bar{V})/\psi$ is independent of β , and so the expression $\beta = \mu(\bar{V})/\psi$ has a unique solution. ■

Lemma 7 *Suppose $r = 0$. Then the reservation quality of buyers is given by the unique solution to:*

$$R + R \sqrt{\frac{m/c - R}{R}} \sqrt{\frac{(\delta + \psi)(1 - \gamma)}{\alpha}} + \gamma s - (1 - \gamma) \frac{m}{c} = 0.$$

Proof. Rewrite equation (18) as

$$s + RF(R) = \int_R^{\bar{q}} 1 - F(\tilde{q}) d\tilde{q} + \alpha (1 - \gamma) \int_R^{\bar{q}} \frac{1 - F(\tilde{q})}{\psi + \delta + \alpha \gamma (1 - F(\tilde{q}))} d\tilde{q}$$

and equation (19) as:

$$1 - F(\tilde{q}) = -\frac{\delta + \psi}{\alpha\gamma} + \frac{1}{\alpha\gamma} \sqrt{\alpha(\delta + \psi)(1 - \gamma)} \sqrt{\frac{(m/c) - q}{q}}.$$

Substituting the second expression into the first yields

$$\begin{aligned}
\alpha\gamma(RF(R) + s) &= \int_R^{\bar{q}} \left[\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q} - \delta - \psi \right. \\
&\quad \left. + \frac{\alpha(1 - \gamma)(\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q} - \delta - \psi)}{\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q}} \right] dq \\
&= \int_R^{\bar{q}} \left[\frac{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q}{\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)}} \right. \\
&\quad \left. + \frac{\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q}[\alpha(1 - \gamma) - \delta - \psi] - \alpha(1 - \gamma)(\delta + \psi)}{\sqrt{\alpha(\delta + \psi)(1 - \gamma)(m/c - q)/q}} \right] dq
\end{aligned}$$

Let I denote the right-hand side of the above equation. To determine the value of I it is useful to perform the following change of variables: $x = \sqrt{\alpha(1 - \gamma)(\delta + \psi)\frac{m/c - q}{q}}$ which implies that

$$q = \frac{\alpha(1 - \gamma)(\delta + \psi)}{x^2 + \alpha(1 - \gamma)(\delta + \psi)} \frac{m}{c}$$

and hence

$$dq = -\frac{2x}{[x^2 + \alpha(1 - \gamma)(\delta + \psi)]^2} \alpha(1 - \gamma)(\delta + \psi) \frac{m}{c} dx,$$

and, moreover,

$$\begin{aligned}
\bar{x} &= \sqrt{\alpha(1 - \gamma)(\delta + \psi)\frac{m/c - \bar{q}}{\bar{q}}} = \delta + \psi \\
\underline{x} &= \sqrt{\alpha(1 - \gamma)(\delta + \psi)\frac{m/c - R}{R}}.
\end{aligned}$$

Therefore we have

$$\begin{aligned}
I &= \int_{\bar{x}}^{\underline{x}} \frac{x^2 + x[\alpha(1-\gamma) - \delta - \psi] - \alpha(1-\gamma)(\delta + \psi)}{x} \left(\frac{2x}{[x^2 + \alpha(1-\gamma)(\delta + \psi)]^2} \alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \right) dx \\
&= 2\alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \int_{\bar{x}}^{\underline{x}} \frac{x^2 + x[\alpha(1-\gamma) - \delta - \psi] - \alpha(1-\gamma)(\delta + \psi)}{[x^2 + \alpha(1-\gamma)(\delta + \psi)]^2} dx \\
&= 2\alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \left[\frac{-\alpha(1-\gamma) + \delta + \psi - 2x}{2(\alpha(1-\gamma)(\delta + \psi) + x^2)} \right]_{\bar{x}}^{\underline{x}} \\
&= 2\alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \\
&\quad \times \left[\frac{-\alpha(1-\gamma) + \delta + \psi - 2\sqrt{\alpha(1-\gamma)(\delta + \psi) \frac{m/c-R}{R}}}{2(\alpha(1-\gamma)(\delta + \psi) + \alpha(1-\gamma)(\delta + \psi) \frac{m/c-R}{R})} - \frac{-\alpha(1-\gamma) + \delta + \psi - 2(\delta + \psi)}{2(\alpha(1-\gamma)(\delta + \psi) + (\delta + \psi)^2)} \right] \\
&= 2\alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \\
&\quad \times \left[\frac{-\alpha(1-\gamma) + \delta + \psi - 2\sqrt{\alpha(1-\gamma)(\delta + \psi) \frac{m/c-R}{R}}}{2\alpha(1-\gamma)(\delta + \psi) \frac{m/c}{R}} - \frac{-[\alpha(1-\gamma) + \delta + \psi]}{2(\delta + \psi)[\alpha(1-\gamma) + (\delta + \psi)]} \right] \\
&= 2\alpha(1-\gamma)(\delta + \psi) \frac{m}{c} \\
&\quad \times \left[\frac{-\alpha(1-\gamma) + \delta + \psi - 2\sqrt{\alpha(1-\gamma)(\delta + \psi) \frac{m/c-R}{R}}}{2\alpha(1-\gamma)(\delta + \psi) \frac{m/c}{R}} + \frac{1}{2(\delta + \psi)} \right] \\
&= R \left[-\alpha(1-\gamma) + \delta + \psi - 2\sqrt{\alpha(1-\gamma)(\delta + \psi) \frac{m/c-R}{R}} \right] + \alpha(1-\gamma) \frac{m}{c}
\end{aligned}$$

This expression for I needs to equal $\alpha\gamma [RF(R) + s]$, which can be rewritten as

$$\begin{aligned}
&\alpha\gamma \left[R \left(1 + \frac{\delta + \psi}{\alpha\gamma} - \frac{1}{\alpha\gamma} \sqrt{\alpha(\delta + \psi)(1-\gamma)} \sqrt{\frac{(m/c) - R}{R}} \right) + s \right] \\
&= \left[R(\alpha\gamma + \delta + \psi) - R\sqrt{\alpha(\delta + \psi)(1-\gamma)} \sqrt{\frac{(m/c) - R}{R}} + \alpha\gamma s \right].
\end{aligned}$$

Equating these two expression yields, after some algebra,

$$R + R\sqrt{(1-\gamma) \frac{(\delta + \psi)}{\alpha}} \sqrt{\frac{m/c - R}{R}} - (1-\gamma) \frac{m}{c} + \gamma s = 0.$$

■

The previous lemma is very useful because we can now compute how the number of buyers depends on the various parameters.

Proof of Proposition 7.

Proof. Noting that $\beta^* = \mu(\bar{V})/\psi$ and $\psi\bar{V} = \alpha(R - m)$ implies that we need only calculate how R depends with each of the variables above. From Lemma 7 we have

$$T(R, s, \delta, \psi, \alpha, c) \equiv R + R\sqrt{\frac{m/c - R}{R}}\sqrt{\frac{(\delta + \psi)(1 - \gamma)}{\alpha}} + \gamma s - (1 - \gamma)\frac{m}{c}$$

and, from the implicit function theorem, we have $\frac{dR}{dy} = -\frac{T_y}{T_R}$ for $y \in \{s, \delta, \psi, \alpha, c\}$.

First, we show that $T_R > 0$:

$$\begin{aligned} T_R &= 1 + \sqrt{\frac{(\delta + \psi)(1 - \gamma)}{\alpha}} \left[\sqrt{\frac{m/c - R}{R}} - \frac{1}{2} \sqrt{\frac{R}{m/c - R}} \frac{m/c}{R} \right] > 0 \\ &\Leftrightarrow \sqrt{\frac{\alpha}{(\delta + \psi)(1 - \gamma)}} \sqrt{\frac{(m/c) - R}{R}} + \frac{m/c}{R} - \frac{m/c}{2R} - 1 > 0 \end{aligned}$$

which resulted from dividing the whole expression by $\sqrt{\frac{(\delta + \psi)(1 - \gamma)}{\alpha}} \sqrt{\frac{R}{(m/c) - R}}$. The sum of the second and third term is positive. The first and fourth can be rearranged as follows:

$$\frac{\alpha}{(\delta + \psi)(1 - \gamma)} \frac{(m/c) - R}{R} > 1$$

Since the left-hand side is decreasing in R , if the inequality is verified for $R = \bar{q}$ then it is verified for all $R \leq \bar{q}$. Replace R by \bar{q} and rearrange to get the inequality $1 > 1 - \gamma$ which holds. Therefore $T_R > 0$.

Looking at the definition of T , it is easy to see that

$$\begin{aligned} T_s > 0 &\Rightarrow \frac{dR}{d\delta} < 0 \\ T_\delta = T_\psi > 0 &\Rightarrow \frac{dR}{d\delta} < 0 \\ T_\alpha < 0 &\Rightarrow \frac{dR}{d\alpha} > 0 \end{aligned}$$

It is claimed that

$$T_{(m/c)} < 0 \Rightarrow \frac{dR}{d(m/c)} > 0.$$

Compute

$$\begin{aligned} T_{(m/c)} &= \sqrt{\frac{(\delta + \psi)(1 - \gamma)}{\alpha}} \frac{1}{2} \sqrt{\frac{R}{(m/c) - R}} - (1 - \gamma) \\ &= \frac{1 - \gamma}{2} \left(\sqrt{\frac{(\delta + \psi)}{\alpha(1 - \gamma)}} \sqrt{\frac{R}{(m/c) - R}} - 2 \right) \end{aligned}$$

To show $T_{(m/c)} < 0$ we need

$$\begin{aligned} 4 &> \frac{(\delta + \psi)}{\alpha(1 - \gamma)} \frac{R}{m/c - R} \\ 4 \frac{m/c - R}{R} &> \frac{(\delta + \psi)}{\alpha(1 - \gamma)} \end{aligned}$$

Using the fact that $R < \bar{q} = \frac{m\alpha(1-\gamma)}{c(\delta+\psi+\alpha(1-\gamma))}$ we have that

$$LHS/4 = \frac{m/c - R}{R} > \frac{m/c - \bar{q}}{\bar{q}} = \frac{\delta + \psi + \alpha(1 - \gamma) - \alpha(1 - \gamma)}{\alpha(1 - \gamma)} = \frac{\delta + \psi}{\alpha(1 - \gamma)} = RHS$$

which proves the claim. ■