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

Manolis Galenianos
Pennsylvania State University

Rosalie Liccardo Pacula
RAND Corporation

Nicola Persico
New York University

August 2008

**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. The model produces testable implications regarding the distribution of purity offered in equilibrium, and the duration of the relationships between buyers and sellers. These predictions are consistent with available data. The effectiveness of different enforcement strategies is evaluated, including some novel ones which leverage the moral hazard present in the market.

*We 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, Iourii Manovski, Jeffrey Miron, Chris Pissarides, Tom Sargent, Jose Scheinkman, Robert Shimer, Gianluca Violante,Travis Wendel, and Randy Wright 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. 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).

# 1 Introduction

The market for illicit drugs is seen as 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.[1] 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 without serious consideration of the retail market structure. The prevailing conception of the retail drugs market is, by default, a Walrasian one: a centralized market with the usual demand and supply curves, and a market-clearing price. In this paper we show that the Walrasian paradigm fails to capture a number of empirical stylized facts, and propose another model which does. The aim of this exercise is not merely descriptive; the model suggests reasons why some current policy interventions may not be effective, and it also suggests new channels for effectively interfering with the retail market.[2]

Our model builds on three basic facts. The first is that retail transactions for illegal drugs are subject to significant moral hazard. What we mean is that the seller can covertly dilute ("cut") the product, and this dilution is largely unobservable to buyers until after they consume. The following table, which is based on data from undercover Drug Enforcement Administration (DEA) purchases, shows that moral hazard is indeed present in this market. 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. 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.[3,4]

[1] In the period 1981-2003 the prison population convicted of drug-related crimes has shot up 9-fold (see Caulkins and Chandler 2006).

[2] Some readers might favor legalization, and thus argue that we should not interfere with the market. Such readers might want to take a *positive,* i.e., descriptive, view of this paper's contributions.

[3] The qualitative results are unchanged if we define a rip-off in terms of quantity*purity rather than in terms of purity only. There is still a significant amount of rip-offs. See Section 5.1.

[4] The 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 ("AT&T," "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 purity | percentage of all trades that are rip-offs (i.e., $\leq 2\%$ purity) | average price of rip-offs (std. dev. of price) | average price of non rip-offs (std. dev. of price) |
|---|---|---|---|---|
| Heroin | 31% | 10.3% | \$53 (22.8) | \$57 (20.6) |
| Crack Cocaine | 68% | 7.8% | \$32 (21.3) | \$38 (24.6) |
| Powder Cocaine | 54% | 5.1% | \$35 (21.8) | \$53 (25.8) |

Table 1: Purity of trades with value $\leq$ \$100 in 1983 dollars.[5]

If this opportunistic behavior is possible, why is it not more prevalent? And, indeed, why doesn't moral hazard foreclose the possibility of trade? The answer lies in the possibility of long-term relationships between buyers and sellers. A seller who wants to keep a customer will not rip him off. Long-term relationships are a key feature of the drugs market and the second basic fact that our model needs to capture.[6,7]

The third basic fact is the presence of considerable dispersion in the price/quality ratio (see Section 5.2). Theoretical models used to evaluate these markets, therefore, need to consider mechanisms that generate this sort of dispersion. In a Walrasian market, we would expect very little variation in the pure grams per dollar spent of a particular substance. In our model, the presence of search frictions and moral hazard generate the equilibrium dispersion of price/quality ratios.

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

[5] Prices computed in 1983 dollars. The number of observations is 12,716 for heroin, 16,202 for crack cocaine, and 5,362 for powder cocaine. These figures are computed from STRIDE, a data collection of undercover purchases by the DEA. These data are discussed briefly in Appendix A.

[6] Buyers derive an additional advantage from long-term relationships: reducing the risk of being caught by undercover police. This risk is captured in our model by a search cost, and it will play an important part in our analysis. See Hoffer (2005) for an interesting ethnographic study of buyers who, over time, manage to hook up with a seller with whom they develop a long-term relationship. We report some systematic evidence of the prevalence of long-term relationships in Section 5.3.

[7] Not 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.

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. We modify the standard Burdett-Mortensen setup by assuming that buyers can only determine the quality of drugs *after* the trade is consummated. This moral hazard leads to severe quality problems, which put the market at risk of collapse; indeed, trade remains possible only because of long-term relationships between buyers and sellers. Introducing moral hazard takes us a long way towards accounting for a number of key stylized facts, such as: a mass of sellers cheat their customers by providing zero-purity drugs; and the wide dispersion in the price/quality ratio, which moreover exhibits a declining density.[8]

The point of this theoretical exercise is three-fold. First, it provides a more realistic description of how the retail markets for illicit drugs operate. The model incorporates a number of frictions, including moral hazard and search costs, which are very salient in the sale and purchase of drugs. These frictions generate a number of testable predictions which match the stylized facts of the retail drugs market and are not accounted for by existing Walrasian models.

Second, the model can help us evaluate 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.[9] Our model presents 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 expands the possibility of trade. These findings highlight the need for an accurate model of market structure in order to evaluate policy.

Third, 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.

[8] On a technical note; it is the presence of moral hazard that generates the required shape of the quality distribution: the Burdett-Mortensen model exhibits an increasing density and no mass points.

[9] The 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).

The rest of the paper is organized as follows. In the next section we provide a brief literature review. In Section 2 we introduce the model; its assumptions are briefly discussed in Section 2.1. Section 3 defines the equilibrium notion we employ. 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 concludes.

## 1.1 Related Literature

Theoretical modeling of drug markets has largely 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 doing business" for the dealer.[10] Our model is considerably more nuanced both in terms of the market structure and in the modeling of enforcement.

Becker *et al.* (2006) assumes a Walrasian market with perfectly informed agents. Enforcement in this model only impacts the market in terms of raising the unit cost of production and/or the full price faced by consumers (in terms of legal risk faced by enforcing laws against users). Their analysis supports the notion 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, this argument works to the extent that the government has the power to tax "legalized" drug sellers without driving them underground. Realistically, even in a world with legalized drugs there would have to be considerable enforcement of "illegal" (i.e., tax-evading) drug sellers, and so our analysis would still apply to that segment of the market.[11]

The problem with the prevailing assumptions of a centralized market and perfectly informed agents is that they abstract from the very elements of drug markets that make them so interesting and they ignore two relevant avenues through which law enforcement can influence the market: by increasing search time and influencing the distribution of purity in the market. Reuter and Caulkins (2004) represents a commendable exception, in that they document the

[10] Reuter and Kleiman (1986) first demonstrated that the risk of being caught is a large component of the cost to the seller. The implication is that enforcement plays a large role in the market price of drugs.

[11] In response to high duties on tobacco in the U.K., a large underground market developed reaching 20% of total tobacco expenditure between 1995 and 1999; see Cullum and Pissarides (2004) p. 12.

large price and quality dispersion in the drugs market,[12] 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.

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).

# 2 The Model

The model highlights three key characteristic of the retail market for illegal drugs: its vulnerability to moral hazard; the fact that many transactions involve a trusted, or regular dealer, who is presumably much less likely to sell diluted drugs; and search costs. In this section we develop a dynamic model that captures these salient features. The focus of the analysis will be to determine what level of quality will be traded for a given amount of money, that is, how high can quality be in equilibrium. A number of the modeling assumptions will be discussed in Section 2.1.

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 1 and a continuum of sellers (or suppliers) of measure $\sigma$. A free entry condition with entry cost $K$ determines the mass of sellers $\sigma$ 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 $\alpha$. 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 $\alpha$, be able to obtain $m$ dollars, which they immediately use to purchase drugs. For simplicity, $m$ is exogenously given and is the same for all consumers.

In return for $m$, the buyer receives $q$. We will refer to $q$ as *quality*. $q$ represents an aggregator

[12] On this point, see also Caulkins *et al.* (2004, 2006).

of quantity and purity, and it captures the utility that the buyer receives from consuming.[13] 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. 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.[14]

The market is characterized by search frictions in the sense that there is no central marketplace where all agents can meet to trade. Rather, buyers and sellers have to trade bilaterally. 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 $\gamma$ that the regular supplier is unavailable, in which case the buyer has to search at random and incur cost $s$. The transition between these two states 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 decides whether to switch sellers if his regular supplier was unavailable and he sampled a new seller at random. A match between a buyer and a seller is exogenously destroyed at rate $\delta$. In this event, the buyer becomes unmatched. The substantive implication of this discussion is that a buyer can only be matched with one seller at a time and there is no recall: once a match has been broken, the buyer cannot find that seller again.

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 effective

[13] Following the criminology literature, we restrict attention to *pure quantity* =quantity*purity as a measure of quality. This assumption is discussed in Section 2.1.

[14] Our equilibrium is consistent with a seller supplying different quality levels to different customers, or with a seller supplying the same quality level to *every customer*. The substantive assumption is that, once a quality has been chosen, it remains fixed for the remainder of the buyer-seller relationship. This 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. We continue our discussion as if a seller sells the same quality to all his customers.

price paid by the buyer—the terms of trade, as it were—and can be thought to capture the *affordability* of drugs. In most of this paper (except possibly for the results in Section 6) we will not need to take a stand on where $m$ comes from. Nevertheless, in the next section we discuss one way to think about the process that determines $m$.

## 2.1 Brief discussion of modeling assumptions

We view our model as most accurately describing the retail, street-level size transactions. It is in those transactions that the moral hazard problem, which is central to the model, is most likely to be important. For 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 absent. We will focus on quality defined as *pure quantity*, where we define pure quantity as the product of (raw) quantity times purity. So, for example, if a transaction involved 0.5 grams of 30% pure heroin, then the pure quantity in that transaction would be 0.5*0.3 = 0.15.

Who is a seller? In our model the seller is a rather abstract agent, who buys wholesale and sells retail. There is nothing in the model that identifies our sellers with single pushers as opposed to criminal gangs. Our model is also silent on the location of purchase—it could be in the street, or in an apartment. What is important for us is that the sellers choose the quality of their product in order to maximize profits. Also, while we model the seller as having the ability to personally cut the drugs, an alternative intepretation 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).

We assume that the buyer can be matched to at most one seller (his "regular" seller). A more sophisticated, and perhaps more realistic assumption could be to allow the buyer some recall of sellers—two or three, say—with whom he was matched in the past. Such sellers might be contacted when the buyer cannot find his regular seller, and the buyer would be forced to incur the search cost $s$ only when all three sellers are unavailable. This more complicated model would not give qualitatively different predictions from the simpler model we analyze.

Quality traded and consumer behavior (which we call "market performance") are unaffected by the size of the mass of sellers (i.e., the buyer/seller ratio). This is shown formally in Proposition 2, but the intuition for this result is straightforward: if, say, a number of sellers exits the market, then the remaining sellers pick up the slack and simply expand their business. In the steady-state equilibrium, buyers are unaffected.[15] By the same token, market performance is unaffected if we allow the mass of consumers to be endogenous—for

[15] In our paper this "irrelevance result" is a direct result of the Burdett-Mortensen matching process we adopt. A different, somewhat related effect was identified in a theoretical paper by Caulkins *et al.* (2006).

example, consumers might be tempted to exit the market if they expect to receive very low purity, or they fear being caught, etc. None of this endogenous entry and exit matters for the equilibrium quality traded or for consumer behavior. This stark feature of the model conveniently allows us to address market performance without having to take a stand on the process that determines demand size.

In our model, $m$ is exogenous—an inelastic amount of money which enters the system at every instant. When money $m$ arrives, consumers spend all of that money, taking whatever quality $q$ that money fetches. No alternative use can be made of $m$, nor can $m$ be increased in order to obtain more $q$. On the other hand, we allow $m$ to be any number. Therefore, our analysis is properly understood as determining the *terms of trade* between $m$ and $q$—for any $m$, how much quality $Q(m)$ that money can fetch. But how is $m$ determined? In reality, consumers—even addicts—have a choice of how much money to devote to drugs consumption. In this bigger scheme, $m$ may be chosen optimally by the consumer trading off her opportunity cost for money against the function $Q(m)$ that this paper provides. After the optimal $m^*$ is determined, our analysis applies directly.

Consumers periodically get the urge/money to consume in our model. The available evidence from the ADAM data set suggests that these urges are pretty frequent for many consumers.[16] 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. The 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 number of times users report consuming the drugs in a month.

Search frictions are introduced in our model for good reason. In the highly decentralized market for illicit drugs, sellers produce and distribute their goods underground, and buyers have to expend resources trying to locate each other without attracting police attention. Sellers cannot broadcast advertising signals that promote their location or the quality of their products. In fact, many of these search frictions are created by law enforcement at a great societal cost. Among the search frictions in our model are exogenous permanent and temporary break-ups of dealer-buyer matches. Permanent break-ups 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. Temporary break-ups are also common. Among the ADAM respondents who

In that paper, it was assumed that jailing the drug dealers would result in more availability of drugs because the jailed dealers would be replaced by even more efficient dealers.

[16] Appendix A has a description of the ADAM data set.

responded to detailed questions about heroin purchases, about a quarter report not being able to purchase heroin in the past 30 days. In many cases, the causes they mention appear temporary in nature (e.g., "police activity," and "no dealer available").[17] If these obstacles can prevent buyers from buying heroin, presumably they also drive buyers to temporarily experiment with new sellers.

# 3 Defining the Equilibrium

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 The 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 $[0, \overline{q}]$. 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:

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

$$\begin{aligned} r\ V(q) = \alpha\ [(1-\gamma)\ q + \gamma\ \int_0^{\overline{q}} (-s + \tilde{q} + \max\{V(\tilde{q}) - V(q), 0\})\ dF(\tilde{q}) - m] \\ + \delta\ (\bar{V} - V(q)). \end{aligned} \tag{2}$$

The interpretation is as follows. Consider equation (1) first. At rate $\alpha$ the buyer gets the

[17] For example, among the 276 respondents who report being unable to buy heroin in the past 30 days, the reasons given are "no dealer available" (33%), "police activity" (28%), "dealer did not have any" (19%), "dealer did not have quality" (14%), "other" (14%). Many of these reasons can be modeled as temporary break-ups.

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 $\alpha$ the buyer wants to consume. With probability $1-\gamma$ his supplier is available and the buyer receives quality $q$. With probability $\gamma$ 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 $m$. Last, at rate $\delta$ 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.[18] 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 $F$.

**Lemma 1** *We have*

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

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

**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 $R = 0$ and $\bar{V} < V(0)$.

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

$$r\,\bar{V} = \alpha\,[-s + \int_0^{\bar{q}} \tilde{q}\, dF(\tilde{q}) + \int_R^{\bar{q}} (V(\tilde{q}) - \bar{V})\, dF(\tilde{q})] \tag{4}$$

$$r\,V(q) = \alpha\,[(1-\gamma)q + \gamma[-s + \int_0^{\bar{q}} \tilde{q}\, dF(\tilde{q}) + \int_q^{\bar{q}} (V(\tilde{q}) - V(q))dF(\tilde{q})] + \delta\,(\bar{V} - V(q)). \tag{5}$$

[18] By assumption, he matches when indifferent.

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

$$R + s = \int_0^{\overline{q}} \tilde{q}\ dF(\tilde{q}) + \int_R^{\overline{q}} (V(\tilde{q}) - \bar{V})\ dF(\tilde{q}) \tag{6}$$

The next step is to integrate by parts the second integral on the right hand side. We start by calculating $V'(q)$. Differentiate equation (5) with respect to $q$ to get

$$rV'(q) = \alpha\ [1 - \gamma + \gamma[-V(q)\ F'(q) - V'(q)\ (1 - F(q)) + V(q)\ F'(q)]] - \delta\ V'(q),$$

whence

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

Now, let us integrate by parts.

$$\begin{aligned}
\int_R^{\overline{q}} (V(\tilde{q}) - \bar{V})\ dF(\tilde{q}) &= \int_R^{\overline{q}} V(\tilde{q})\ dF(\tilde{q}) - (1 - F(R))\bar{V} \\
&= V(\overline{q}) - V(R)\ F(R) - \int_R^{\overline{q}} F(\tilde{q})\ V'(\tilde{q})\ d\tilde{q} - (1 - F(R))V(R) \\
&= \int_R^{\overline{q}} (1 - F(\tilde{q}))\ V'(\tilde{q})\ d\tilde{q},
\end{aligned}$$

where in the last step we used $V(\overline{q}) - V(R) = \int_R^{\overline{q}} V'(\tilde{q})\ d\tilde{q}$. Substituting the results of this integration into equation (6) yields equation (3). ■

## 3.2 The Sellers' Decision Problem

The seller's problem is to choose a level of quality that maximizes his steady-state level of profits. The steady state profits of a seller who chooses to offer quality level $q$ are given by

$$\pi(q) = (m - c\ q)\ t(q) \tag{7}$$

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

The assumption that sellers maximize their steady state level of profits allows us to side-step the possible complexities due to short-term fluctuations in the number of customers. This fluctuation is due to the randomness involved in the matching technology, which means that two sellers offering the same quality might have different number of buyers. In the long-run, or if there are very many buyers per seller, this randomness averages out giving rise to an expected number of buyers $t(q)$. In this case there is no complexity to worry about. But if sellers are impatient *and* the number of buyers per seller is low, then the short-run randomness would matter for the sellers' profits. In that case, we would have to keep track the number of buyers per *each* seller as a state variable. Expression (7) side-steps this complexity by focusing on the steady state.

### 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 $\sigma$ 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: the mass of sellers $\sigma$ is such that the profit level equals the cost of entry, $\bar{\pi} = K$.*

## 4 Characterization of Equilibrium

### 4.1 Existence and Uniqueness

Equilibria in our model exist and are unique. Depending on parameter values, the distribution of quality traded $q$ may exhibit a mass point at zero—a feature of particular interest for us. The following proposition states these results formally.

**Theorem 1** *The equilibrium exists and is unique. In equilibrium, a mass point of sellers offers zero quality if and only if $c \cdot s < k$, where $k > 0$ is a function of parameters.*

**Proof.** See Appendix B. ■

Theorem 1 says that the equilibrium is more likely to exhibit a mass of sellers offering zero quality, if search costs $s$ are low. Intuitively, this is because when $s$ is small, buyers are picky about which seller to match with, which in turn makes it less rewarding for sellers to offer high quality. 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, the equilibria when there is no mass point at zero are characterized in Appendix B.

## 4.2 Steady-State Flow of Trades

In this section we characterize the equilibrium level of sales $t(q)$ of a seller who offers quality $q$, as a function of the distribution of quality offered $F$, and of the buyers' reservation quality $R$. The only assumption used in this section is that we are at a steady state, and so the aggregate population statistics (number of matched and unmatched sellers, etc.) are constant over time.

Sales come from two sources: the steady state number of "loyal" customers, and the random customers who sample once and may or may not become regular after consuming (if they do become regulars, they are counted as 'loyal' from then on). Denote the flow of loyal and occasional buyers by $t_L(q)$ and $t_O$, respectively. The flow of total sales is given by

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

We now solve for $t(q)$.

**Lemma 2** *The steady state level of transactions of a seller offering quality $q$ is given by*

$$t(q) = \frac{\alpha}{\sigma} \; \frac{\delta + \alpha\ \gamma\ (1 - F(R))}{\delta + \alpha\ (1 - F(R))} \; [1 + \frac{\alpha\ \delta\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1 - F(q))]^2}], \quad \textit{when } q \geq R, \tag{8}$$

$$t(q) = \frac{\alpha}{\sigma} \; \frac{\delta + \alpha\ \gamma\ (1 - F(R))}{\delta + \alpha\ (1 - F(R))}, \quad \textit{when } q < R. \tag{9}$$

**Proof.** When deriving the transaction flow it will prove more convenient to work with the number of buyers rather than sellers, so let $\beta = 1/\sigma$ denote the total number of buyers per seller. Let the steady state number of unmatched and matched buyers be given by $n$ and $\beta - n$, respectively. In steady state, the flows of buyers from the matched state to the unmatched state and vice versa equal each other. 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 becomes unmatched when his match is exogenously destroyed which occurs at rate $\delta$. As a result, in steady state the following holds:

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

Isolating $n$ yields

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

The flow of occasional customers consists of unmatched buyers and of matched buyers whose regular supplier is unavailable. Therefore

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

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 need to solve for $l(q)$. It is immediate that $l(q) = 0$ when $q < R$. To describe $l(q)$ for $q \geq R$, let $G$ denote the distribution of qualities that matched buyers receive. $G$ first order stochastically dominates the distribution of offered qualities because a matched buyer moves to higher qualities over time. The number of matched buyers receiving quality up to $q$ is given by $(\beta - n)\ G(q)$. The flow into this group comes from the $n$ 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 the flows and solving for $G(q)$ yields

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

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

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

The average number of buyers matched to a seller offering quality level $q$ is given by

$$l(q) = \lim_{\epsilon \to 0} (\beta - n) \, \frac{G(q) - G(q-\epsilon)}{F(q) - F(q-\epsilon)} = (\beta - n) \, \frac{G'(q)}{F'(q)}$$

(we assume here, and later verify, that $F$ is differentiable for $q > R$). It is a matter of algebra to arrive at

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

Substituting into (10) yields the result. ■

## 4.3 Characterizing the Quality Distribution

To characterize the quality distribution $F$, we use the fact that in equilibrium all qualities that are offered yield the same steady state profits. Increasing the quality offered affects profits in two ways: it reduces the margin per transaction and it increases the number of transactions by raising the steady state number of regular customers. The number of customers increases in $q$ 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 $F$.

**Proposition 1** *For any $m$, the following properties hold in equilibrium:*
*(i) If $q$ is offered by a seller then either $q = 0$ or $q \geq R$.*
*(ii) $F$ has no mass point on the positive part of its support.*
*(iii) $F$ exhibits quality dispersion.*
*(iv) The positive part of the support of $F$ is connected and is given by $[R, \overline{q}]$.*
*(v) $\overline{q} = \frac{m}{c} \cdot \frac{\alpha \; \delta \; (1-\gamma)}{\delta + \alpha \; (1-\gamma)}$.*
*(vi) On the positive part of its support, $F$ is given by*

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

**Proof.** (i) A seller who offers $q \in [0, R)$ has no regular customers. As result $t(q) = t_O$ for all $q \in [0, R)$ and any positive quality is dominated by $q = 0$.

(ii) 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.

(iii) The argument above proves that there will be quality dispersion unless every seller offers 0. Suppose that $F(0) = 1$. Then it is easy to see that lemma 1 implies $R = 0$ and the no-mass point argument yields a contradiction.

(iv) Suppose there is a gap in the support of $F$ between $q_1$ and $q_2$, where $R \leq q_1 < q_2 \leq \overline{q}$. The sellers offering $q_1$ and $q_2$ have exactly the same number of regular 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.

(vi) 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 $\mathbf{R}_+$):

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

On the interval $[R, \overline{q}]$, 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,

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

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

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

■

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 ($\varepsilon$ can be very small).

### 4.4 Solving for the mass of sellers

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

**Proposition 2** *1. The equilibrium quality distribution $F$ does not depend on the mass of sellers $\sigma$.*

*2. Profits per firm are multiplicative in $1/\sigma$.*

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

**Proof.** 1. Expression (11) does not depend on $\sigma$.

2. Observe from equations (8) and (9) that $1/\sigma$ enters profits multiplicatively, scaling the total number of transactions $t(q)$. Furthermore, $\sigma$ 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 $\sigma$ decreases. If, as $\sigma$ decreases, the remaining sellers keep offering a quality distribution according to $F$, then condition (7) is preserved. This means that it is an equilibrium for the remaining sellers to offer quality distribution $F$ and for buyers to leave their strategy unchanged. After the change, every seller's profits are scaled up by the same amount.

3. Immediate from part 2. ■

This proposition completes the description of equilibrium.

## 5 Testable Implications

The model has implications both for the cross-sectional distribution of qualities traded and for the time series of individual consumption. In this section we work out three sets of implications for which there is empirical evidence to compare to. The first set of implications concerns the qualitative features of the cross-sectional quality distribution. As the empirical counterpart to quality $q$ we take *pure quantity* (the product of quantity times purity). Our model can match the qualitative features of the cross-sectional distribution of pure quantity, as observed in the STRIDE data set.[19]

The second set of implications has to do with consumer loyalty. First, in our model consumers are expected to be somewhat loyal to their seller. Moreover, the model predicts that consumers who trade more frequently are also more picky: they will not settle for a low-quality seller, but rather they will keep searching until they find a high quality one. When these consumers do settle down with a regular seller, that seller will really be high-quality and thus these consumers are less likely to abandon him in favor of another seller. This means that frequent consumers, once they settle down, are more loyal to their seller. The ADAM data set provides evidence both of long term relationships and of greater loyalty by frequent consumers.[20]

The third set of implications concerns the correlation between wholesale price and retail affordability.

## 5.1 Quality Rip-Offs

Our theoretical analysis has focussed on parameter constellations such that the equilibrium exhibits a positive mass of rip-offs. In this section we construct an empirical counterpart to rip-offs. We label as rip-offs those trades that yield a *pure quantity* which is less than 10% of the average pure quantity traded. The next table indicates that rip-offs are present in all three types of drugs.

| Drug | average pure quantity in grams | percentage of all trades that are rip-offs $\leq 0.01g$ for crack and cocaine $\leq 0.016g$ for heroin | average price of rip-offs (std. dev. of price) | average price of non rip-offs (std. dev. of price) |
|---|---|---|---|---|
| Heroin | 0.16 | 15.4% | \$47.1 (22.8) | \$58 (20.1) |
| Crack Cocaine | 0.46 | 7.9% | \$31.5 (21.3) | \$37.7 (24.6) |
| Powder Cocaine | 0.64 | 5.2% | \$34.3 (21.6) | \$53.2 (25.8) |

Table 2: Pure quantity of trades with value $\leq$ \$100 in 1983 dollars.[21]

[19] The STRIDE data set is discussed briefly in Appendix A.

[20] The ADAM data set is discussed briefly in Appendix A.

[21] Prices computed in 1983 dollars. The number of observations is 12,716 for heroin, 16,202 for crack cocaine, and 5,362 for powder cocaine.

Even the relatively high incidence of rip-offs found in Table 2 may be an underestimate of the mass point in the distribution $F$ offered by sellers. 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.[22]

## 5.2 Quality Dispersion

We next discuss the shape of the quality distribution offered in equilibrium.

**Proposition 3** *The density distribution of sellers offering quality $q$ is monotonically decreasing in $q$ if (and only if)* $\frac{\alpha\,(1-\gamma)}{\delta+\alpha\,(1-\gamma)} \leq \frac{3}{4}$.

**Proof.** Differentiating expression (11) yields the density of the qualities offered in equilibrium,

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

The function $f\left(q\right)$ is a strictly decreasing transformation of $\left(\frac{m}{c}-q\right)q^3$. The latter function has a unique (local and global) maximum at $q = \frac{3}{4}\,\frac{m}{c}$. Therefore, $f\left(q\right)$ has a unique minimum at $q = \frac{3}{4}\,\frac{m}{c}$. The support of $F$, remember, has upper bound $\overline{q}$. So if $\overline{q} > \frac{3}{4}\,\frac{m}{c}$, 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. ■

Figure 1 shows a qualitative plot of the density distribution of the quality offered by sellers in equilibrium, under the assumption that $\frac{\alpha\,(1-\gamma)}{\delta+\alpha\,(1-\gamma)} \leq \frac{3}{4}$.

We would like to compare this theoretical distribution with the empirical distribution of pure quantity. As an example, Figure 2 depicts the distribution of pure quantity of crack cocaine traded for \$20 in Washington, DC in the period 1989-1991. This empirical distribution resembles the theoretical distribution displayed in Figure 1.

Figure 2 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 2 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

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

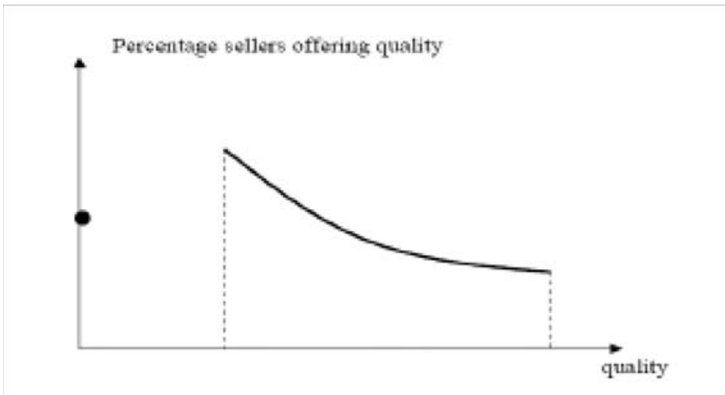


Figure 1: Quality offered in equilibrium

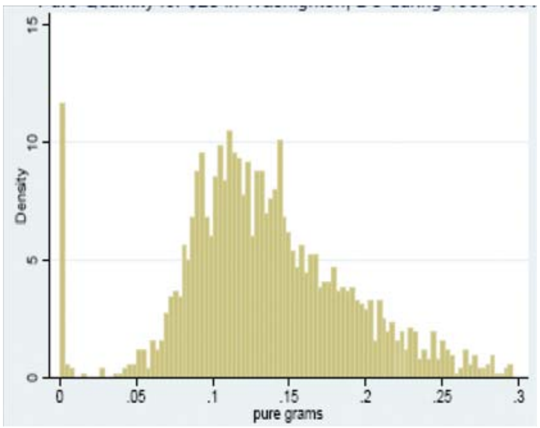


Figure 2: Pure quantity of crack traded for $20 in Washington DC, 1989-1991.

to devise a strategy for aggregating many pictures like Figure 2. The first limitation is addressed by studying pure grams *per hundred 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 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. If we take Figure 1 to represent the quality distribution in a city/year, then the normalization we adopt rescales the horizontal axis but it does not change the shape of the distribution. (The normalization causes each of these city/year distributions to be centered around 1.) We then aggregate the normalized observations by drugs type, and display the results in Figure 3.[23]

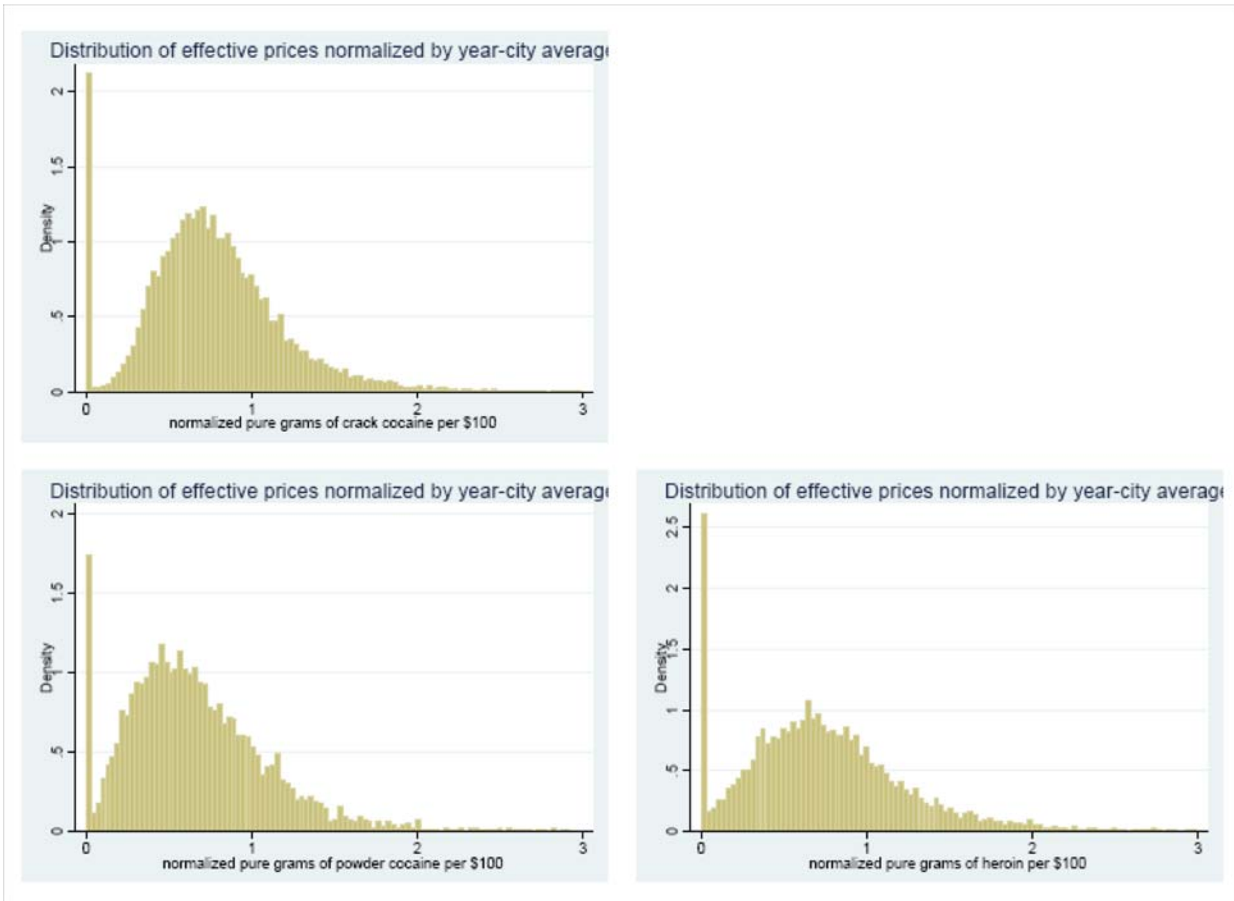


Figure 3: Normalized pure grams per $100.

As might be expected, the empirical distributions in Figure 3 are less sharp, more smooth, than the theoretical prediction in Figure 1. We ascribe this lack of sharpness to aggregation issues. Aside from this difference, several points of similarity are worth noting between the empirical and theoretical distributions. First, they both feature a relatively large mass point at zero quality. Second, they both represent a gap above zero quality, although the gap in the empirical data is a bit fuzzier as there are some observations that do fall into the range before the modal response in the positive purity part of the distribution. Third, both show a decreasing density in the quality distribution after the modal response in the positive range of the purity distribution. Finally, both demonstrate an enormous degree of quality

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

dispersion.

The fourth point, the degree of quality dispersion, is a very robust feature of the data. In Appendix C, we provide evidence that this dispersion is not an artefact of aggregation across time and space. Therefore, the presence of this dispersion can be interpreted as a violation of the "law of one price." Clearly, the Walrasian model is not a good metaphor for the retail drugs market. In our view, this massive quality dispersion represents strong evidence in favor of a model with search frictions, such as the one presented in this paper.

On the whole, we think that the qualitative features of the cross-sectional distribution of quality predicted by the model are consistent with the available empirical evidence.

## 5.3 Long-term Relationships and Loyalty

Repeated transactions are an integral part of the way our model operates. The next table, compiled from the ADAM data set, provides (buyer-reported) evidence of a large amount of repeat business. The presence of repeat business is, of course, consistent with the equilibrium of our model. The table also provides further detail. It indicates that more frequent consumers appear to be more loyal to their regular suppliers. The table below shows that for heroin, for example, 76% of frequent users obtained their last purchase from their regular supplier, while only 58% of casual users did.

| | Heroin | | Crack Cocaine | | Powder Cocaine | |
|---|---|---|---|---|---|---|
| Last supplier | 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 3: Repeated transactions.[24]

Our model would yield a similar prediction if we introduced buyer heterogeneity in the frequency of consumption. The following lemma proves a key partial equilibrium result which speaks to this issue. Given *any* distribution of quality $F$ (not necessarily an equilibrium one), a buyer who consumes more often (has a higher $\alpha$) has a higher reservation quality.

**Lemma 3** *Given any distribution of quality, a buyer who consumes more often has a higher reservation quality: formally, $\partial R/\partial\alpha > 0$.*

[24]For each drug, these are the male respondents who reported consuming that drug at least once in the previous 30 days.

**Proof.** Inspecting equation (3) shows that the right-hand side monotonically increases in $\alpha$. Therefore, as $\alpha$ increases an increase in $\hat{R}$ is required to preserve equality in equation (3). ■

This lemma shows that, 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. This result has direct implications for the equilibrium of a more elaborate version of our model, one in which buyers are heterogeneous in their $\alpha$'s. While we do not explicitly set up and solve such a model, we note the property stated in Lemma 3 holds for any $F$. Thus, whatever $F$ might arise in the equilibrium of a model with heterogeneous $\alpha$'s, the reservation level of the more frequent consumers will be higher. This argument suggests that in a model with heterogeneity in the frequency of consumption, the more frequent buyers will be "more loyal," consistent with the data.

## 5.4 Effects of Wholesale Price Changes

What happens to affordability (the distribution of qualities offered $F$ per money $m$) when the wholesale cost $c$ changes? It is possible to give sharp comparative statics results concerning the median of $F$.

**Proposition 4** *Suppose less 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.** Let $\phi$ 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 $\tilde{F}(q) = \max\left[\tilde{\mathbb{F}}(q), \tilde{\mathbb{F}}(\tilde{R})\right]$; the strict inequality comes from $\partial\mathbb{F}(q)/\partial c > 0$ (cf. equation 11); and the second-to-last equality follows from $y > F(0)$. ■

Figure 4 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.[25] The figure shows two patterns. First, wholesale (mean) price and retail median quality are negatively correlated,

[25] "Retail" transactions are those worth less than \$100, and "wholesale" transactions are those worth more than \$1,000, in 1983 dollars.

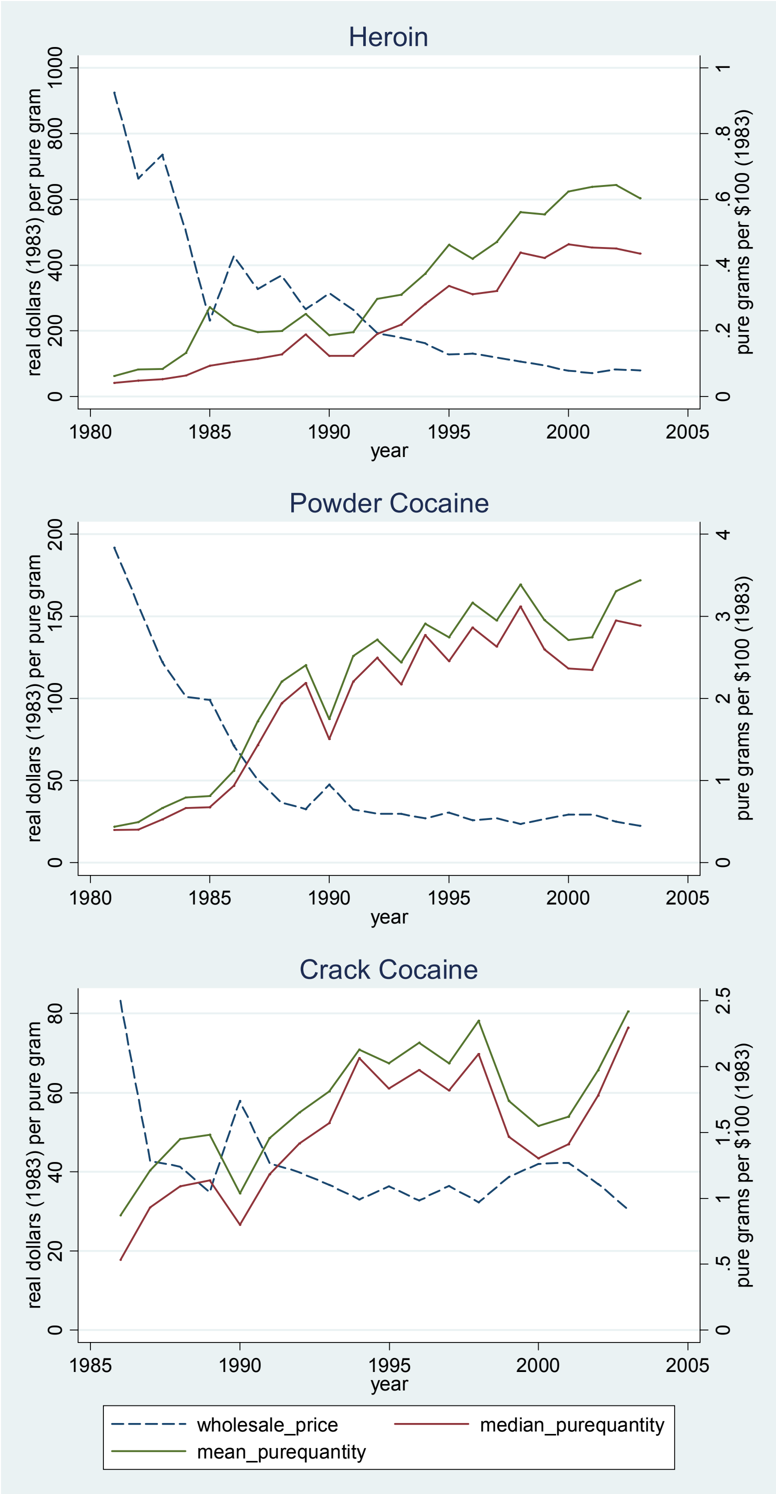


Figure 4: Wholesale price, and retail quality (mean and median).

consistent with the theoretical prediction of Proposition 4. Second, Figure 4 also shows a very similar correlation between the wholesale price and retail *average* quality. This empirical correlation, although not implied by Proposition 4, is consistent with our numerical simulations.

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.

# 6 Comparative Statics

In this section we provide 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 unity of money $m$ spent).

A foreword on the interpretation of the comparative statics results. If we want to think about how to use enforcement and interdiction to reduce quality traded, we may want to bear in mind that in reality $m$ is potentially endogenous. If $m$ 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 $\widetilde{m}^*$ 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.[26]

## 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 several effects of increased enforcement.[27]

We find that simply deterring some sellers need not *per se* affect the affordability of drugs. 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 5 Part 1).

[26] 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.

[27] The comparative statics in points 1 and 2 can also capture changes in sentencing.

We also find, counterintuitively, that increasing the search cost $s$ results in improved quality of drugs (Proposition 5 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 ($\gamma$ in our model). Proposition 5 Part 3 indicates that increasing the temporary separation rate need not reduce drug affordability.

**Proposition 5** ***(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 $\gamma$ increases, the average and median quality of drugs offered by sellers for given $m$ (affordability) may increase.*

**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 $\gamma$. 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. ■

Proposition 5 paints a complex picture of the effects of enforcement. Increasing enforcement on sellers achieves deterrence, but deterrence has no effect on the affordability of drugs. Reducing $s$, the buyer's search cost, reduces affordability. However, the effect on buyer surplus is ambiguous because reducing $s$ has a direct effect of increasing buyer surplus (buyers who search incur cost $s$ in equilibrium). Thus reducing $s$ might induce consumer entry into the market. Similarly, increasing $\gamma$, the temporary separation rate is also not necessarily advisable. First, Proposition 5 does not give a monotonic prediction, so we may increase $\gamma$ 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 affordability and market participation in complex and potentially counterintuitive 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.[28] Formally, we study the following penalty scheme:

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

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 $\overline{q}$ 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.[29] Given a mass 1 of undercover officers who meet with sellers at a constant Poisson rate $\zeta$, the arrest rate for an individual seller is equal to $\zeta/\sigma$. 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 6** ***(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 the same effect on affordability as increasing $c$, see Proposition 4), 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 4); and (b) it also decreases the jail population.*

[28] In practice, introducing discounts for low purity might drive sellers to work on a margin not studied in this paper, i.e., increase the amount of cutting agent per trade without decreasing quality per unit of money $m$. 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.

[29] These undercover officer exist: they are the ones who collect much of the data in STRIDE.

**Proof.** 1. Suppose the penalty is as in (P1). The number of people entering jail in each instant is given by $\zeta$ and the amount of time they spend in jail is $J$, so the steady state prison population is $\zeta\ J$. 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 $\zeta$, 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:

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

where we denote $\widetilde{m} = m - \frac{\zeta}{\sigma}\ J$. We know from Proposition 1 part (vi) that the equilibrium $F$ depends solely on the ratio $\widetilde{m}/c$. 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

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

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

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

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

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

where we denote $\widehat{m} = m - \frac{\zeta}{\sigma}\ (J - j\ \bar{q})$ and $\widehat{c} = c + \frac{\zeta}{\sigma}\ j$. We know from Proposition 1 part (vi) that the equilibrium $F$ depends solely on the ratio $\widehat{m}/\widehat{c}$. Increasing $j$ decreases this ratio provided that $\widehat{m} > \overline{q}\widehat{c}$, 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 $c$. ■

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 Conclusions

Over the last 25 years the "war on drugs" has channelled enormous amounts of resources into interfering with the drugs market, yet the theoretical understanding of market structure has not evolved beyond the most stark Walrasian models. In this paper we have provided a relatively sophisticated theoretical model of the market structure of the retail drugs trade. The model 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 "scam" transactions, the shape of the quality distribution, and the patterns of consumer loyalty.

The model is quite unlike a Walrasian model, and for good reason. The available evidence strongly rejects the "law of one price." Rather, the large "price dispersion" found in the data speaks in favor of a model with search frictions, such as the one studied here.

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. Perhaps these results can also shed light on the limited success of some components of enforcement policies.

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.

# 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] Becker GS and KM Murphy. 1988. "A Theory of Rational Addiction" *Journal of Political Economy* 96: 675-700.

[3] 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.

[4] Berensten, A., and G. Rocheteau (2004). "Money and information". *Review of Economic Studies* 71 (October), pp. 915–44.

[5] Burdett, Kenneth, and Dale T. Mortensen (1998). "Wage Differentials, Employer Size, and Unemployment." *International Economic Review,* pp. 257-273.

[6] Burdett, Kenneth, Ricardo Lagos, and Randall Wright (2003). "Crime, Inequality, and Unemployment." *American Economic Review* 93(5), pp. 1764-1777.

[7] 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.*

[8] 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.

[9] 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.

[10] 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.

[11] Coles, Melvyn (2001). "Equilibrium wage dispersion, firm size, and growth".*Review of Economic Dynamics* 4(1), pp. 159-187 January 2001.

[12] Cullum, Paul, and Christopher A. Pissarides (2004) "The Demand for Tobacco Products in the UK." Government Economic Service Working Paper No 150, December 2004.

[13] Engelhardt, Bryan, Guillaume Rocheteau, and Peter Rupert (2008). "Crime and the labor market: A search model with optimal contracts." Forthcoming, J*ournal of Public Economics.*

[14] 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.

[15] 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.

[16] Hoffer, Lee D. (2005) *Junkie Business: The Evolution and Operation of a Heroin Dealing Network* Wadsworth Publishing; 1st edition (December 13, 2005).

[17] 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).

[18] Reuter P and J Caulkins (2004) "Illegal Lemons: Price Dispersion in the Cocaine and Heroin Markets" *UN Bulletin on Narcotics.*

[19] 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.

[20] 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

[21] Schelling T 1984. "Self-command in practice, in policy, and in a theory of rational choice." *American Economic Review* 74. pp. 1-11.

[22] Stigler GJ and GS Becker. 1977. "De Gustibus Non Est Disputandum" *American Economic Review* 67, pp. 76-90.

[23] Trejos, Alberto (1999). "Search, Bargaining, Money, and Prices Under Private Information." *International Economic Review* 40(3), pp. 679-696

[24] 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.

[25] 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.

# A The data

We concentrate our attention to 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 has a total of approximately 780,000 observations for a number of different drugs and acquisition methods. The STRIDE data 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.[30] 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 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 ADAM data set is collected quarterly from interviews with arrestees in 35 counties across the country. 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. 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 for all arrestees who report having used drugs in the previous 30 days. Information collected includes number of times drugs was 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.

# B Proof of Theorem 1.

We now provide the full characterization of equilibrium for all possible parameter values.

**Proposition 7** *Equilibria exist and they are unique. All equilibria exhibit quality dispersion. The quality distribution has a continuous part above the buyers' reservation, $R$.*

[30] The latter is determined through chemical analysis in a DEA laboratory

*The buyers' reservation, $R$, and the number of sellers offering zero quality, $F_0$, are determined as follows:*

*There are cutoffs $\underline{s}$ and $\overline{s}$ such that*

$$\begin{aligned} &\textit{If } s < \underline{s}, \textit{ then } F_0 = \hat{F}_0 \textit{ and } R = \hat{R}, \\ &\textit{If } s \in [\underline{s}, \overline{s}], \textit{ then } F_0 = 0 \textit{ and } R = \hat{R} \\ &\textit{If } s > \overline{s}, \textit{ then } F_0 = 0 \textit{ and } R = 0. \end{aligned}$$

*where $\hat{R}$ and $\hat{F}_0$ are strictly positive and they are functions of the parameters.*

We shall prove the proposition in a series of lemmata. Existence and uniqueness are proved constructively.

First, we fix some arbitrary values for the buyers' reservation and the number of sellers offering zero quality where the only restrictions are $R \geq 0$ and $F_0 \in [0, 1)$. Given some arbitrary pair $R,\ F_0$ we characterize the distribution of offered qualities such that all sellers in the positive part of the distribution earn the same steady state profits.

Second, we solve for the buyers' reservation as a function of the, still arbitrary, number of sellers who offer zero quality and the, previously derived, positive part of the quality distribution.

Third, we determine how many sellers, if any, offer zero quality by requiring that offering zero yields the same steady state profits as offering $R$.

**Lemma 4** *Given some $F_0 \in [0, 1)$ and $R > 0$, the distribution of offered qualities is*

$$F(q) = \begin{array}{ll} 1 + \frac{\delta}{\alpha\ \gamma}(1 - \sqrt{h(q)}) & \textit{for } q \in [R, \overline{q}) \\ F_0 & \textit{for } q \in [0, R], \end{array} \tag{13}$$

*where*

$$h(q) = \frac{m/c - q}{\frac{\delta\ (q-R)}{\alpha\ (1-\gamma)} + \frac{m/c-R}{[1+\alpha\ \gamma\ (1-F_0)/\delta]^2}} \tag{14}$$

$$\overline{q} = [\frac{m}{c} + \frac{\delta\ R}{\alpha\ (1-\gamma)} - \frac{m/c - R}{[1 + \alpha\ \gamma\ (1 - F_0)/\delta]^2}]/(1 + \frac{\delta}{\alpha\ (1-\gamma)}) \tag{15}$$

**Proof.** We construct $F$ so that the profits from any positive quality level on offer are the same.

Recalling that $\pi(q) = t(q)\ (m - c\ q)$ and using the derivation of $t(q)$ from lemma 2 yields that the profits of a seller offering $R$ and some $q \geq R$ are:

$$\pi(R) = \alpha\ \frac{\delta + \alpha\ \gamma\ (1 - F_0)}{\delta + \alpha\ (1 - F_0)}\ [1 + \frac{\alpha\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1 - F_0)]^2}]\ (p - c\ R)$$
$$\pi(q) = \alpha\ \frac{\delta + \alpha\ \gamma\ (1 - F_0)}{\delta + \alpha\ (1 - F_0)}\ [1 + \frac{\alpha\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1 - F(q))]^2}]\ (p - c\ q)$$

Equating the two expressions and solving for $F$ gives (13). Setting $F(\overline{q}) = 1$ and solving for $\overline{q}$ yields (15). ■

**Lemma 5** *Given some $F_0 \in [0, 1)$, the buyers' reservation is*

$$R = \hat{R} \quad \textit{if } s < \overline{s}$$
$$R = 0 \quad \textit{if } s \geq \overline{s}$$

*where*

$$\overline{s} = \frac{m\ \alpha\ (1-\gamma)}{c\ \delta}\ [\frac{\alpha\ \gamma\ (1 - F_0)}{\delta + \alpha\ \gamma\ (1 - F_0)}]^2 \tag{16}$$
$$\hat{R} = \frac{m\ H(F_0)/c - s}{F_0 + H(F_0)}, \tag{17}$$
$$H(F_0) = \alpha\ (1-\gamma)/\delta\ [\frac{\alpha\ \gamma\ (1 - F_0)}{\delta + \alpha\ \gamma\ (1 - F_0)}]^2 \tag{18}$$

**Proof.** We fix some $F_0 \in [0, 1)$. We combine equation (3), which gives $R$ as a function of $F$, with (13), which gives the positive part of $F$ as a function of $R$ and $F_0$. The purpose is to derive $R$ as a function of $F_0$ alone.

We first derive $\hat{R}$ and then check whether it is greater than zero. Since $F(q) = F_0\ \forall\ q \leq R$,

$$\int_0^{\overline{q}} \tilde{q}\ dF(\tilde{q}) = \ (1 - F_0)\ R + \int_R^{\overline{q}} (1 - F(\tilde{q}))\ d\tilde{q}$$

which is a direct result of integration by parts. Hence we can rewrite equation (3) as

$$\hat{R}\ F_0 = -s + \int_{\hat{R}}^{\overline{q}} (1 - F(\tilde{q}))\ [1 + \frac{\alpha\ (1-\gamma)}{\delta + \alpha\ \gamma\ (1 - F(q))}]\ dq.$$

Using (13), this expression can be simplified to

$$\begin{aligned}\hat{R}\ F_0 &= -s + \int_{\hat{R}}^{\overline{q}} \frac{\delta}{\alpha\ \gamma}\ [\sqrt{h(q)} - 1]\ [1 + \frac{\alpha\ (1-\gamma)}{\delta\ \sqrt{h(q)}}]\ dq \\ &= -s + \frac{\delta}{\alpha\ \gamma}\ \{[\frac{\alpha\ (1-\gamma)}{\delta} - 1]\ (\overline{q} - \hat{R}) + \int_{\hat{R}}^{\overline{q}} \sqrt{h(q)} - \frac{\alpha\ (1-\gamma)}{\delta\ \sqrt{h(q)}}\}\ dq \qquad (19)\end{aligned}$$

To integrate, let

$$\begin{aligned}a_1 &= m/c > \overline{q} \\ a_2 &= \frac{\delta}{\alpha\ (1-\gamma)} \\ a_3 &= -\frac{\delta\ \hat{R}}{\alpha\ \gamma} + \frac{\delta\ (m/c - \hat{R})}{[\delta + (1-F_0)]^2}\end{aligned}$$

Then equation (14) implies that $h(q) = \frac{a_1 - q}{q\ a_2 + a_3}$ where $a_1, a_2, a_3 > 0$ and the key is to evaluate the following:

$$I = \int_{\hat{R}}^{\overline{q}} [\sqrt{h(q)} - \frac{1}{a_2\ \sqrt{h(q)}}]\ dq \qquad (20)$$

To evaluate this integral we perform a change of variables. Define $w = h(q)$, i.e. $w = \frac{a_1 - q}{a_2\ q + a_3}$. This implies that $q = h^{-1}(w) = \frac{a_1 - a_3\ w}{a_2\ w + 1}$ and $dq = (h^{-1}(q))' dw = -\frac{a_1\ a_2 + a_3}{(a_2\ w+1)^2}\ dw$. Furthermore, the limits of integration become $h(\overline{q})$ and $h(\hat{R})$. As a result, we can rewrite the integral as

$$\begin{aligned}I &= \int_{h(\hat{R})}^{h(\overline{q})} [\sqrt{w} - \frac{1}{a_2\ \sqrt{w}}]\ (-\frac{a_1\ a_2 + a_3}{(a_2\ w + 1)^2})\ dw \\ &= (a_1\ a_2 + a_3) \int_{h(\overline{q})}^{h(\hat{R})} [\frac{\sqrt{w}}{(a_2\ w+1)^2} - \frac{1}{a_2\ \sqrt{w}\ (a_2\ w + 1)^2}]\ dw\end{aligned}$$

Next, let $z = \sqrt{w}$ so that $w = z^2$, $dw = 2\ z\ dz$ and $q = \hat{R} \Rightarrow z = \sqrt{h(\hat{R})}$, $q = \overline{q} \Rightarrow z = \sqrt{h(\overline{q})}$. The integral becomes

$$I = (a_1\ a_2 + a_3) \int_{\sqrt{h(\overline{q})}}^{\sqrt{h(\hat{R})}} [\frac{z}{(a_2\ z^2 + 1)^2} - \frac{1}{a_2\ z\ (a_2\ z^2 + 1)^2}]\ 2\ z\ dz$$
$$= 2\ (a_1\ a_2 + a_3) \int_{\sqrt{h(\overline{q})}}^{\sqrt{h(\hat{R})}} [\frac{z^2}{(a_2\ z^2 + 1)^2} - \frac{1}{a_2\ (a_2\ z^2 + 1)^2}]\ dz$$

The last change of variables is to let $\theta = \arctan(\sqrt{a_2}\ z)$, so that $z = \frac{\tan\theta}{\sqrt{a_2}}, dz = \frac{1}{\sqrt{a_2}\ \cos^2\theta}\ d\theta$ and $q = \hat{R} \Rightarrow \theta = \arctan(\sqrt{a_2\ h(\hat{R})}) \equiv \theta(\hat{R})$, $q = \overline{q} \Rightarrow z = \arctan(\sqrt{a_2\ h(\overline{q})}) \equiv \theta(\overline{q})$. Note that $a_2\ z^2 + 1 = \tan^2\theta + 1 = \frac{1}{\cos^2\theta}$ since $\tan\theta = \sin\theta/\cos\theta$ and $\sin^2\theta + \cos^2\theta = 1$. Therefore, we can rewrite $I$ as

$$I = 2\ (a_1\ a_2 + a_3) \int_{\theta(\overline{q})}^{\theta(\hat{R})} [\frac{\tan^2\theta/a_2}{1/\cos^4\theta} - \frac{1}{a_2/\cos^4\theta}]\ \frac{1}{\sqrt{a_2}\ \cos^2\theta} d\theta$$
$$= \frac{2\ (a_1\ a_2 + a_3)}{(a_2)^{3/2}} \int_{\theta(\overline{q})}^{\theta(\hat{R})} [\sin^2\theta - \cos^2\theta] d\theta$$
$$= \frac{2\ (a_1\ a_2 + a_3)}{(a_2)^{3/2}} \int_{\theta(\overline{q})}^{\theta(\hat{R})} [2\ \sin^2\theta - 1] d\theta$$
$$= \frac{2\ (a_1\ a_2 + a_3)}{(a_2)^{3/2}} [\cos\theta(\overline{q})\ \sin\theta(\overline{q}) - \cos\theta(\hat{R})\ \sin\theta(\hat{R})]$$
$$= \frac{2\ (a_1\ a_2 + a_3)}{(a_2)^{3/2}} [\frac{\tan\theta(\overline{q})}{1 + \tan^2\theta(\overline{q})} - \frac{\tan\theta(\hat{R})}{1 + \tan^2\theta(\hat{R})}]$$

where we have used the trigonometric identities $\sin^2\theta + \cos^2\theta = 1$, $\cos^2\theta = \frac{1}{1+\tan^2\theta}$ and $\sin^2\theta = \frac{\tan^2\theta}{1+\tan^2\theta}$.

To complete the evaluation of (20), observe that equation (13) implies that $\theta(\overline{q}) = \sqrt{h(\overline{q})} = 1$ and $\theta(\hat{R}) = \sqrt{h(\hat{R})} = 1 + \alpha\ \gamma/\delta\ (1 - F_0)$. Putting things together and going through the algebra yields

$$I = 2\ \alpha\ (1-\gamma)\ (\frac{m}{c} - \hat{R})\ (\frac{1}{\delta} + \frac{\alpha\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1 - F_0)]^2})\ (\frac{\delta}{\alpha\ (1-\gamma) + \delta}$$
$$- \frac{\delta + \alpha\ \gamma\ (1 - F_0)}{\alpha\ (1-\gamma) + \delta\ [1 + \alpha\ \gamma/\delta\ (1 - F_0)]^2})$$

While not particularly attractive, the above equation can be entered into (19) to arrive at the expression for $\hat{R}$ in equations (17) and (18). Recalling that $R = \max\{\hat{R}, 0\}$ yields

$$R = \hat{R} \Leftrightarrow$$
$$s \leq \frac{m\ \alpha\ (1-\gamma)}{c\ \delta}\ [\frac{\alpha\ \gamma\ (1-F_0)}{\delta + \alpha\ \gamma\ (1-F_0)}]^2$$

which yields (16). ■

**Lemma 6** *The number of sellers offering zero quality is given by*

$$F_0 = \hat{F}_0 \quad \textit{if } s \leq \underline{s} \tag{21}$$
$$F_0 = 0 \quad \textit{if } s > \underline{s} \tag{22}$$

*where*

$$\hat{F}_0 = 1 + \frac{\xi_2 + \sqrt{\xi_2^2 - 4\ \xi_1\ \xi_3}}{2\ \xi_1} \tag{23}$$
$$\xi_1 = [\frac{\delta}{\alpha\ (1-\gamma)} - \frac{m}{c\ s}]\ (\frac{\alpha\ \gamma}{\delta})^2,\ \ \xi_2 = \frac{2\ \alpha\ \gamma}{\alpha\ (1-\gamma)} - \frac{m}{c\ s},\ \ \xi_3 = 1 + \frac{m}{c\ s} + \frac{\delta}{\alpha\ (1-\gamma)} \tag{24}$$

**Proof.** To determine $F_0$ we use the condition that zero- and positive-quality sellers make the same steady state profits in equilibrium.

We compare the profits of zero-quality sellers with those offering $R$. The steady state profits to a seller who offers 0 and $R$ are given by

$$\pi(0) = \alpha\ \frac{\delta + \alpha\ \gamma\ (1-F_0)}{\delta + \alpha\ (1-F_0)}\ m \tag{25}$$
$$\pi(R) = \alpha\ \frac{\delta + \alpha\ \gamma\ (1-F_0)}{\delta + \alpha\ (1-F_0)}\ [1 + \frac{\alpha\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1-F_0)]^2}]\ (m - c\ R) \tag{26}$$

which leads to the following expression:

$$\Delta\pi(F_0) \equiv \pi(R) - \pi(0) \tag{27}$$
$$= \alpha\ \frac{\delta + \alpha\ \gamma\ (1-F_0)}{\delta + \alpha\ (1-F_0)}\ [(m - c\ R)\ \frac{\alpha\ (1-\gamma)}{\delta\ [1 + \alpha\ \gamma/\delta\ (1-F_0)]^2} - c\ R] \tag{28}$$

In equilibrium, $\Delta\pi(F_0) = 0$ if $F_0 > 0$. Note that if $R = 0$ (i.e. if $\hat{R}(F_0) \leq 0$) then $\Delta\pi(F_0) > 0$ which means that non-opportunistic sellers enjoy higher profits than opportunistic ones. This can only be in equilibrium if no one actually offers 0, i.e. if $F_0 = 0$. Therefore, from now on we introduce $\hat{R}$ in (27), knowing that if $\hat{R} \leq 0$ then $F_0 = 0$ which is the correct answer.

A bit of algebra leads to

$$\Delta\pi(F_0) = \frac{\alpha^2\ \delta\ c\ s\ (1-\gamma)}{[\delta + \alpha\ \gamma\ (1-F_0)]\ [\delta + \alpha\ (1-F_0)]\ [F_0 + H(F_0)]}\ N(F_0), \tag{29}$$

where

$$N(F_0) = \frac{m\ F_0}{c\ s} + 1 - \frac{m\ [\alpha\ \gamma/\delta\ (1-F_0)]^2}{c\ s} + \frac{[\delta + \alpha\ \gamma\ (1-F_0)]^2}{\delta\ \alpha\ (1-\gamma)} \tag{30}$$

The first term of (29) is always positive. Therefore, $\Delta\pi(F_0) = 0$ for some $F_0$ only if $N(F_0) = 0$. Letting $\bar{F}_0 \equiv 1 - F_0$ we can rewrite (30) as

$$N(\bar{F}_0) = \xi_1\ \bar{F}_0^2 + \xi_2\ \bar{F}_0 + \xi_3 \tag{31}$$

where the $\xi$s are defined in (24).

The fact that $\xi_3 > 0$ means that $N(0) > 0$ which confirms that in equilibrium positive quality is offered by a strictly positive measure of sellers. If $\xi_1 \geq 0$, then $\hat{R} \leq 0$ which leads to $F_0 = 0$. Concentrating on the case where $\xi_1 < 0$, we can find the roots of (31). The positive root gives $\bar{F}_0 < 0$. The negative root is in $[0, 1]$ when $N(1) < 0$ which is equivalent to $s \leq \overline{s}$. ■

This completes the characterization of equilibria.

# C Quality Dispersion

Below we explore quality dispersion in our sample. For example, in our entire heroin sample, \$100 buys on average 0.39 pure grams of heroin with a very large standard deviation of 0.58, which leads to a coefficient of variation (standard deviation over mean) of 1.48. 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, the time variable is the year; in the second, it is the quarter that the transaction took place. Before conducting the fixed effects regression we restrict our sample to cities that have more than 400 observations in total in the case of the year, and 950 observations in the case of the quarter regressions. Furthermore, we drop all transactions that belong to city-time cells with fewer than 5 observations. In the first case we are left with 19,072 data points from 23 cities and in the second with 12,955 data points from 8 cities. We then compute the coefficient of variation, using the standard deviation of the residuals divided by the sample mean without fixed effects.

**Heroin**

| Sample | N | mean | std dev | coeff of variation |
|---|---|---|---|---|
| Full | 29,187 | 0.39 | 0.58 | 1.48 |
| Restricted 1 | 19,072 | 0.38 | 0.55 | 1.45 |
| Restricted 1 - city/year/interaction fixed effects | 19,072 | 0.00 | 0.47 | 1.24 |
| Restricted 2 | 12,195 | 0.39 | 0.53 | 1.36 |
| Restricted 2 - city/quarter/interaction fixed effects | 12,195 | 0.00 | 0.46 | 1.18 |

**Crack**

| Sample | N | mean | std dev | coeff of variation |
|---|---|---|---|---|
| Full | 47,743 | 1.83 | 1.51 | 0.83 |
| Restricted 1 | 20,262 | 1.73 | 1.49 | 0.86 |
| Restricted 1 - city/year/interaction fixed effects | 20,262 | 0.00 | 1.32 | 0.76 |
| Restricted 2 | 12,743 | 1.45 | 1.38 | 0.95 |
| Restricted 2 - city/quarter/interaction fixed effects | 12,743 | 0.00 | 1.25 | 0.86 |

**Cocaine**

| Sample | N | mean | std dev | coeff of variation |
|---|---|---|---|---|
| Full | 46,050 | 2.12 | 1.98 | 0.93 |
| Restricted 1 | 18,862 | 2.14 | 1.98 | 0.93 |
| Restricted 1 - city/year/interaction fixed effects | 18,862 | 0.00 | 1.59 | 0.74 |
| Restricted 2 | 9,030 | 2.09 | 2.07 | 0.99 |
| Restricted 2 - city/quarter/interaction fixed effects | 9,030 | 0.00 | 1.68 | 0.80 |

| | |
|---|---|
| Restricted 1 | keep cities with more than 400 observations<br>drop if fewer than 5 observations in city-year cell |
| Restricted 2 | keep cities with more than 950 observations<br>drop if fewer than 5 observations in city-quarter cell |

Both restricted samples have very similar summary statistics to the full sample. One remarkable feature of the summary statistics of the residuals is that the standard deviation decreased by only 17% and 15% when adding city/time/interaction fixed effects, even though one of the dominant stylized facts of the drugs markets in the last 25 years is the dramatic decrease in price. Our results suggest that most of the price variation occurs within a point in space and time rather than across different points. Finally in both regressions the coefficient of variation (using the mean prior to the regression which, if anything, yields an underestimate) is still more than 3 times larger than the one for wine.