A Model of Chaotic Drug Markets and Their Control

Doris A. Behrens
Vienna University of Technology

Jonathan P. Caulkins
Carnegie Mellon University

Gustav Feichtinger
Vienna University of Technology

Follow this and additional works at: http://repository.cmu.edu/heinzworks
A Model of Chaotic Drug Markets and Their Control

D.A. BEHRENS, J.P. CAULKINS, G. FEICHTINGER

Working Paper No. 268
June 2002
A Model of Chaotic Drug Markets and Their Control

Doris A. Behrens¹,², Jonathan P. Caulkins³,⁴, and Gustav Feichtinger¹

¹ Department of Operations Research and Systems Theory, Vienna University of Technology, Argentinierstrasse 8, A–1040 Vienna, Austria
² Department of Economics, University of Klagenfurt, Universitaetsstrasse 65-67, A–9020 Klagenfurt, Austria, email: doris.behrens@uni-klu.ac.at
³ H. John Heinz III School of Public Policy Management, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, U.S.A., email: caulkins@andrew.cmu.edu
⁴ RAND, Drug Policy Research Center, 201 North Craig St., Suite 102, Pittsburgh, PA 15213, U.S.A., email: caulkins@rand.org

Abstract: Drug markets are often described informally as being chaotic, and there is a tendency to believe that control efforts can make things worse, not better, at least in some circumstances. This paper explores the idea that such statements might be literally true in a mathematical sense by considering a discrete-time model of populations of drug users and drug sellers for which initiation into either population is a function of relative numbers of both populations. The structure of the system follows that considered in an arms control context by Behrens et al. (1997). In this context, the model suggests that depending on the market parameter values, the uncontrolled system may or may not be chaotic. Static application of either treatment or enforcement applied to a system that is not initially chaotic can make it chaotic and vice versa, but even if static control would create chaos, dynamic controls can be crafted that avoid it. So-called OGY controls seem to work well for this example.

Keywords: Nonlinear dynamic systems, chaos, OGY control, drug markets, static and dynamic drug control policies.

+ This research was partly financed by the Austrian Science Foundation under contract No. P14060-SOZ.
* Corresponding author: Doris Behrens, Department of Economics, University of Klagenfurt, Universitaetsstrasse 65-67, A–9020 Klagenfurt, Austria, email: doris.behrens@uni-klu.ac.at, phone: +43(463) 2700-4125, fax: +43(463) 2700-4191.
1. Introduction

Illicit drug use and drug-related crime impose significant costs on society. Many of the problems stem more directly from the drug markets than from the drug use per se (MacCoun and Reuter, 2001), and law enforcement directed at market participants within US borders has constituted the bulk of drug control efforts in the US (ONDCP, 2000).

Drug markets, particularly street markets, are often informally described as “chaotic.” E.g., neighbors troubled by street markets complain that they are disorderly, and law enforcement sees the markets as being unpredictable, resilient, and having a life of their own. There is a natural tendency to want the government to do something about such problems, but there are also concerns that heavy-handed or simple-minded control efforts can make things worse. E.g., removing sellers may disturb markets in ways that ultimately exacerbate violence (e.g., Maher and Dixon, 2001). On the other hand, carefully timed and targeted interventions have generated some clear success stories (e.g., Kennedy, 1997).

Statements that drug markets are “chaotic” are often understood to be figurative, but we explore here the possibility that they may be rooted in a more literal, mathematical notion of chaos. We have in mind here public or “street” markets for illicit drugs and the users who might patronize such geographically-defined markets. At the national level, market aggregates do not seem to be chaotic. They vary over time in more modest ways that can be explained by more conventional models. Likewise, there always exist social-network based distribution channels that are difficult for drug enforcement to penetrate. But brazen public markets do seem to appear, grow, shrink, and occasionally disappear in multiple locations throughout most major American cities in ways that follow some regularities (e.g., they are more common in neighborhoods with low social cohesion) but nonetheless are hard to explain or predict.

Because street markets cause so many problems and so many resources are devoted to their control, a considerable literature has developed modeling the impact of so-called “crackdowns” on these markets (Barnett, 1988; Baveja et al., 1993, 2000; Caulkins, 1993; Moore and Kleiman, 1989; Kleiman, 1988; Kort et al., 1998; Weisburd and Green, 1995). The present model is related inasmuch as it addresses street markets, but different in that it focuses on the dynamic interplay between the numbers of users and sellers. Earlier efforts tended to view numbers of sellers and users as changing in concert and so had an essentially one-dimensional characterization of market size. Lee (1993) distinguished between users and sellers, but focused on how best to divide enforcement effort between users and sellers. This paper has a richer model of the uncontrolled dynamics and
contrasts enforcement against sellers with anything that removes users from the population, whether that be enforcement or treatment as in Tragler et al., (2001). One can also draw comparisons with two-population “predator-prey” models from the mathematical biosciences literature dating back to Lotka (1920) and Volterra (1931), but the most direct analog is Behrens et al.’s (1997) study of arms control.

The model presented in this paper summarizes prerequisites for possible chaotic behavior of addicts and drug dealers in an open-air drug market. Interestingly, even if the uncontrolled market dynamics do not exhibit chaotic patterns, a static intervention like removing a constant fraction of addicts each time period can create chaos, quasi-periodic or at least periodic behavior. Furthermore, we show how erratic behavior can be removed by the application of intertemporal (optimal) control.

2. The Descriptive Model

We are interested in modeling the number of users \( (A_t) \) and sellers \( (D_t) \) in a city’s public drug market(s) at time \( t \). For the sake of simplicity, we assume that both populations tend to decay at a constant per capita rate that is counterbalanced by an inflow which is a function of the relative numbers of users and sellers. The inflow of new sellers is highest when there are many more users than sellers simply because such markets offer the most attractive profit opportunities. The initiation of new users is driven primarily by social contact with existing users, not by sellers. So initiation of new users is increasing in the number of existing users and is not affected directly by the number of sellers, as long as the drug is generally available, as is the case for most major drugs in the US. Because of two indirect effects, however, it is quite plausible that initiation is actually decreasing in the number of street sellers. First, street markets generate violence and other problems that can give the drug a negative reputation, thereby dampening initiation. (See Musto (1987) for more on negative feedback effects from reputation.) Second,

---

1 Contrary to once popular belief, drug sellers do not “push” their drugs on unsuspecting people in order to stimulate demand. Kaplan gives a good overview of the reasons (Kaplan, 1983). In economic terms, the inability to subsequently charge the new initiates above market prices makes it unwise to hook them with “loss leaders”. Also, sellers’ greatly increase their risk of enforcement sanction when they expose their illegal activities to people not themselves always involved in the market.

2 The Monitoring the Future study consistently finds that for each of the major illicit drugs more than a third of high school seniors report that the drug is “fairly easy” or “very easy” to obtain, and this high availability has persisted despite substantial increases in enforcement levels. (Johnston, et al., 2001)
large numbers of street sellers might reorient current users into the drug market “scene” and away from their usual social milieu where they frequently interact with non-users and, hence, spread drug use to new users. That is, public markets may be thought of as serving primarily established users, not potential new initiates, and vibrant street markets may create social separation between current users and potential initiates.

One could argue instead that initiation ought to be unaffected by or even be increasing in both the number of street sellers, but we do not pursue those possibilities here both because historically drug initiation fell sharply in the US during the 1980s when street markets were expanding and simply because postulating a negative effect of street sellers on initiation creates symmetry between the impacts of users and sellers on each other’s population flows. Specifically,

\[
A_{t+1} = (1 - \alpha)A_t + a f(A_t, D_t), \quad A_0 = A(0),
\]

\[
D_{t+1} = (1 - \beta)D_t + b f(A_t, D_t), \quad D_0 = D(0),
\]

(1)

with \(a, b > 0\), \(\alpha, \beta \in [0, 1]\),

where \(f(A, D) > 0\) has \(f_A > 0\) and \(f_D < 0\). When users are plentiful relative to street sellers \((A > D)\), the number of users and sellers increases (see Equation (2) and Figure 1) at constant rates, although the effect may be more pronounced for sellers (i.e., plausibly but not necessarily \(b > a\)). For the function \(f(A, D)\) we use

\[
f(A, D) = \frac{1}{1 + e^{-c(A-D)}} \quad \text{with} \quad c > 0
\]

(2)

which can be seen as a continuous approximation to the discontinuous step function. The parameter \(c\) determines whether the approximation is close (large \(c\)) or if there is a smooth transition between markets that tend to grow or shrink (small \(c\)). A sharp transition would seem most plausible if the time step is large and/or there exist good substitutes for these street markets, e.g., established distribution channels embedded in social networks or a tradition of “beeper sales”.

For interpretive convenience, we will think of \(A (D)\) as measuring the proportion of the maximum number of users (sellers) who might ever be active, so that \(A (D)\) remains within the unit interval (where we know from Behrens et al. (1997) that for System (1) and (2) and for \(a \in (0, \alpha]\) and \(b \in (0, \beta]\) \(A_t, D_t \in [0, 1]\) implies \(A_{t+1}, D_{t+1} \in [0, 1]\) for all \(t \geq 0\). This can be done without loss of generality since the system is bounded and we have not yet specified the units of the constants \(a\) or \(b\).
Figure 1: Initiation function $f$ plotted for three different “steepness” parameters $c$.

Figure 2: Flow diagram of Model (1) and (2).

The parameter $\alpha$ in system (1) is the outflow rate for addicts from death, recovery, etc., and $\beta$ is the outflow rate of dealers from death, incarceration, reform, etc. Since addiction careers of 5 to 20 years are not uncommon, but a full-time retail hard drug seller in the US has about a one-third chance of being incarcerated over the course of a year (Reuter et al., 1990) we assume that
and interpret the time step as being one year or less.

Model (1) and (2) provides an example of the interaction of negative and positive feedback. The number of dealers, $D$, and the number of addicts, $A$, have opposite impacts on initiation into $A$ and $D$. It is a common fact that such a combination of stabilizing and destabilizing effects may lead to periodic or quasi-periodic and even chaotic trajectories. Hence, it is necessary to go beyond a linear analysis of stability and use some numerical methods to investigate the system in a global sense. (See, e.g., Khibinik et al. (1993), Medio (1992), or Schaffer et al. (1988).) Chaotic behavior is most likely to emerge when $\alpha < \beta$, $a < b$, and when $c$ is of intermediate magnitude (Behrens, 1992). In particular, $c$ must be smaller than some $c_2 < \infty$ and larger than some $c_1 > 0$, where the values of $c_1$ and $c_2$ depend on the remaining parameters, $\alpha$, $\beta$, $a$, and $b$.

Thus, conditions for chaos can be met by the drug market model above. That is, $\alpha$ is certainly less than $\beta$, $a$ is likely less than $b$, and since $c$ may well vary from city to city or market to market it is not implausible that in some markets $c$ falls within the necessary range. We stress that this is an argument for plausibility. We are by no means saying that for all cities or markets the parameters necessarily fall in a range that produces chaos.

3. The Dynamics of the Modeled Drug Market

In general, turmoil in drug markets is harmful (Maher and Dixon, 2001). Although policy may seek to increase the uncertainty market participants face as a way of driving some out of the market, this is a means to an end. For any given average number of participants, gyrations typically exacerbate externalities. Clearly this is a broad generalization to which exceptions no doubt exist, but in general if it is not possible to have no drug market at all (clearly the best outcome) then a stable drug market may be the best one can hope for. Constant variation over time is bad, and the worst possible situation may be unpredictable change. Among the reasons is that change breeds misunderstanding and conflict among market participants who have no recourse to formal means of dispute resolution (such as the court system), and so too often resolve disputes violently (Reuter, 1983).

3 If $c$ gets really large the function $f$ approximates the discontinuous step function and chaos is removed in most cases. Having either $\alpha \geq \beta$ or $a \geq b$ often seems to remove chaos, but by no means guarantees that chaos does not occur (Behrens et al., 1992).
Accepting for the moment this hierarchy of degrees of market harms, a natural question to ask is whether some intervention or control can ameliorate a drug market’s harms by shifting it from chaotic variation to stable cycles or from cycles to a constant (possibly low level) equilibrium.

With this model it is natural to think of two kinds of interventions: those that eliminate sellers and those that eliminate users. The former is essentially the function of enforcement. The latter could be done through enforcement, but in practice not very many people are incarcerated solely because they are drug users, so we will refer to it as treatment. At least initially we will find it convenient to think of these interventions as operating on the per capita rates of exit from the respective population. I.e., treatment increases the parameter \( \alpha \), and enforcement increases the parameter \( \beta \) (but maybe also moderates the initiation rate \( b \)).

To visualize possible types of system behavior and the effect of these controls, we have to determine parameter values. These parameters are chosen purely for illustrative purposes and do not necessarily reflect any specific drug or location. For one such parameter set,

\[
S = \{\alpha = 0.2977, \beta = 0.4580, a = 0.2000, b = 0.3900, c = 161.1536\},
\]

Figure 3 shows the period-doubling route to chaos (via a cascade of Flip bifurcations\(^4\)). Close to the chaotic range we additionally find solutions with prime periods 7, ..., 2x5, ..., 2x3, and, finally, 3. Note that according to the Sarkovskii Theorem (see, e.g., Strogatz, 1994), the existence of a period 3 orbit implies the existence of orbits of any other period.

Hence, we can easily imagine that the following scenario might occur. Let us assume that a stable drug market exists. I.e., the corresponding system is not chaotic but exhibits monotonous or periodic patterns of evolution. In Figure 3 one such situation might be described by \( \beta < \alpha + 0.14 \). If then a decision-maker chooses to expand enforcement, effectively increasing \( \beta \) (depicted by the vertical arrow in Figure 3) the inequality \( \beta < \alpha + 0.14 \) may be violated. This could direct the system into chaos. Conversely, if the situation were initially in the stable region described by \( \beta > 1.25\alpha + 0.1375 \), and the decision maker expanded treatment (increased \( \alpha \)), one might likewise observe a progression from simple to more complex

\[^4\text{Note that System (1) and (2) exhibits Flip, Fold, Transcritical and Neimark-Sacker bifurcations (Behrens et al. 1997).}\]

\[^5\text{All calculations were done using DYNAMICAL SOFTWARE (Schaffer et al., 1988), DMC (Medio, 1992), LOCBIF (Khabinik et al., 1993), and MATHEMATICA 4.1 (Wolfram, 1996) under the Windows 98 operating system. To verify chaos, e.g., in Figures 3 and 4, we used local bifurcation analysis, investigations of the Lyapunov spectra, and estimates of the fractal dimension.}\]
Figure 3: System behavior for different combinations of desistance rates for $a = 0.2$, $b = 0.39$ and $c = 161.1536$. Initial conditions are given by $A_0 = D_0/2 = 0.13$.

Figure 4: System behavior for different combinations of initial values for $a = 0.2$, $b = 0.39$, $\alpha = 0.2977$, $\beta = 0.458$ and $c = 161.1536$. 
and, ultimately, chaotic dynamics (depicted by the horizontal arrow in Figure 3). In both cases there can be alternating regions of chaotic and regular motion, so-called “periodic windows”. Thus, even within the “chaotic region”, data generated by Model (1) and (2) do not necessarily exhibit erratic patterns but follow orbits of very low periodicity (see Figure 3).

Furthermore, the outcome of such policy changes could depend on their timing. Figure 4 illustrates this for an intervention that pushes the system to $\alpha = 0.2977$ and $\beta = 0.4580$. When users were plentiful relative to sellers, (three-)periodic orbits would result. The same change at a different point in time when sellers were plentiful relative to users would lead to chaos.

Changing the parameter $c$ can also induce chaos. Increasing $c$ from zero to $c_1 \approx 132$ generates a period doubling route to chaos as depicted in Figure 5 for parameter set $S = \{\alpha = 0.2977, \beta = 0.4580, a = 0.2000, b = 0.3900, c = 161.1536\}$. But then if $c$ increases beyond a critical value of $c_2 \approx 180$ the bifurcation diagram shows that chaotic behavior is suddenly replaced by the appearance of a three-periodic attractor. Recall that $c$ reflects the steepness of the “response function” $f$ and likely depends on, among other things, the existence of alternative markets. This suggests that developments in other, related markets (e.g., the beeper sales market) might affect the stability of place-based street markets.

![Figure 5: Bifurcation diagram of parameter $c$ for $a = 0.2$, $b = 0.39$, $\alpha = 0.2799$ and $\beta = 0.458$. Initial conditions are given by $A_0 = D_0/2 = 0.13$.](image-url)
4. Controlling the Market / Removing Chaos

We have seen that static interventions can be counter-productive (compare Behrens et al., 1999, 2000, 2002). One way to avoid chaos would be to implement dynamic interventions. We explore this idea here with a dynamic control that removes users (whether through treatment or incarceration) combined with static policy toward removing sellers, thereby showing that dynamic control of just one of the two levers can be enough to pull a system back out of chaos. We will also slightly generalize the analysis by allowing the static control of selling to not only remove sellers (increase $\beta$) but also suppress their initiation (decrease $b$). In particular, we model the static enforcement policy, represented by the term $v$, as altering the flow rates associated with dealers in system (1) and (2), $0 < a < b$, $c \in (c_1, c_2)$ and $\alpha, \beta \in [0, 1]$, as follows:

$$\tilde{\beta} := \beta + v \quad \text{and} \quad \tilde{b} := b - \gamma v,$$

where

$$\gamma > 0 \quad \text{and} \quad v \in [0, \min\{1 - \beta, b/\gamma\}).$$

If the static enforcement policy $v$ happens to satisfy

$$\alpha - \beta < v < \frac{b - a}{\gamma},$$

in addition to condition (5), then for $a < b, \gamma \in (0, \infty)$, $c \in (c_1, c_2)$, and $\alpha, \beta \in [0, 1]$ chaos will appear with all likelihood in system (1), (2), (4), and (5) (Behrens, 1992). Thus, inequality (6) represents a necessary (but not sufficient) condition for the occurrence of chaos caused by a static enforcement policy. If we merge conditions (5) and (6),

$$v \in \left[\max\{0, \alpha - \beta\}, \min\{1 - \beta, (b - a)/\gamma\}\right],$$

we notice that static enforcement polices applied with “a low intensity” are more likely to “create” unpredictable market behavior than are more intensive (and, hence, expensive) interventions. So a city or region might find itself confronted by the following dilemma. Persisting with very intense enforcement implies high control costs, but easing enforcement could trigger chaos in drug markets that is costly in other ways, e.g., through increased disorder and violence. Figure 6 illustrates that in the absence of control, a market characterized by parameter set $\mathcal{S} = \{\alpha = 0.2977, \beta = 0.4580, a = 0.2000, b = 0.3900, c = 161.1536\}$ never settles down to an equilibrium.
A possible way out of such a dilemma is implementing a complementary, possibly dynamic, treatment intervention policy. A (dynamic) treatment program, $0 \leq u_t \leq 1 - \alpha$, can be modeled by transforming system (1), (2), (4), and (7) into

\[
\begin{align*}
A_{t+1} &= (1 - \alpha - u_t) A_t + a f(A_t, D_t), \quad A_0 = A(0), \\
D_{t+1} &= (1 - \tilde{\beta})D_t + \tilde{b} f(A_t, D_t), \quad D_0 = D(0),
\end{align*}
\]

where $f$ is defined by Equation (2).

In what follows we consider several types of treatment programs distinguished on a rather technical basis (how their intensity is adjusted over time, if at all) not according to what these programs include (e.g., distinguishing therapeutic communities from methadone maintenance).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Chaotic attractor, unstable one-period fixed point and its stable and unstable directions, respectively, of System (1) and (2) for $a = 0.2, b = 0.39, \alpha = 0.2977, \tilde{\beta} = 0.458$ and $c = 161.1536$. $A_0 = D_0/2 = 0.13$ and the time horizon is equal to $t = 50, \ldots, 5,000$.}
\end{figure}
4.1 Static Treatment Policies

The most “convenient” type of intervention one might think of is “constant treatment”, i.e., helping a constant proportion of addicts to desist from drug use per time unit. Unfortunately, we cannot give any general condition for a constant control ruling out the occurrence of chaos. (Recall that having either \( \alpha \geq \beta \) or \( a \geq b \) by no means guarantees that chaos does not occur.) However, it is possible to find a critical threshold for a non-chaotic control by local bifurcation analysis (searching only over a one-dimensional, bounded parameter) and the calculation of Lyapunov spectra using an appropriate software package (see, e.g., Medio (1992) or Schaffer et al. (1988)). In doing so we will pay particular attention to “regions” of constant treatment implying non-periodic behavior. The associated conditions for constant treatment can be determined by investigating the eigenvalues of the linearized system evaluated at the fixed point (see Behrens et al., 1997).

So we face a number of (potential) problems with relying on static control. First, the extinction of chaos cannot be guaranteed. Second, sometimes the level of treatment required to remove chaos approaches \( \bar{u}_c = 1 - \alpha \), which would imply removing almost all current users every period. Treatment programs are simply not that effective; relapse is common. Third, even if such intense control were feasible, it might be prohibitively expensive.

4.2 Dynamic Treatment Policies

A reliable way to get rid of chaos is provided by the Ott-Grebogi-Yorke (OGY) algorithm (Ott et al., 1990; Shinbrot, 1993). By making small time-depended perturbations of the parameter \( \alpha \) (which is basically equivalent to a dynamic treatment policy) the OGY method can stabilize the system to a (formerly unstable) fixed point of system (8). Treatment policies derived by the OGY method, \( u_t^{OGY} \), have to satisfy following condition:

\[
0 \leq u_t^{OGY} \leq 1 - \alpha. \tag{9}
\]

---

6 Given a parameter configuration of the drug market model that yields chaotic system dynamics, the OGY algorithm starts with examining the unstable periodic orbits that are embedded in the chaotic attractor and selects the one which gives the “best system’s performance”. Since stability is our primary aim this procedure can be reduced to calculating the unique (one-periodic) fixed point.

7 Since negative treatment would not make sense, condition (10) differs from the “usual” OGY condition that says the deviation from the base case value \( \alpha \) (which is treatment in our interpretation) has to stay in a tube around the stable manifold.
We use knowledge of how points move near the one-periodic fixed point \( \hat{X}^\alpha \) (Equation A3) under a change of the “control parameter” \( \alpha \) (by adding \( u_t^{OGY} \)) to determine the parameter change needed to move the trajectory to a point located on the stable manifold of the fixed point \( \hat{X}^\alpha \) (see Appendix A1). Once placed (exactly) on the stable manifold the trajectory naturally converges towards the desired fixed point \( \hat{X}^\alpha \) so that the system can be stabilized near the saddle point \( \hat{X}^\alpha \) of the system. We receive following relation for the OGY control derived from Equation (A7) in Appendix A.1,

\[
u_t^{OGY} = -\frac{\lambda_u}{y_{1u}} \left(1 + e^{-c(A_{t} - D_{t})}\right)\left(v_{1u} \left(A_{t} - \hat{A}^\alpha\right) + y_{2u} \left(D_{t} - \hat{D}^\alpha\right)\right)
\]

which can be used as long as condition (9) is satisfied. The scalar \( \lambda_u \) represents the unstable eigenvalue of the fixed point \( X^\alpha = (A^\alpha, D^\alpha)^T \) calculated for \( \alpha \) (where \( T \) indicates the transpose), and \( y_{1u} = (v_{1u}, y_{2u}) \) denotes the contravariant basis vector that is perpendicular to the stable eigenvector of the Jacobian evaluated at \( X^\alpha \).

One drawback of the OGY control is that it can take time before the system gets close enough to the stable manifold for the OGY control to operate. An alternative is to aggressively “steer” the system toward the manifold by a procedure called targeting (Shinbrot et al., 1992a,b) so the OGY control can “begin” sooner. (Figure 6 illustrates this (linearized) stable manifold and denotes with an arrow the idea of steering down to it rather than waiting for the uncontrolled system to wander near it.) For examples of such a combined form of chaos control see Kopel (1997, 1998, 1999).

The main purpose of OGY control with or without steering is to remove chaos and not to reduce some cost function. Thus, it would be of interest to compare the OGY controls with a “cost optimal” dynamic control minimizing the objective function,

\[
J(T) = \sum_{t=0}^{T} \frac{1}{(1+r)^t} \left(E(A_t) + F(D_t) + G(u_t)\right),
\]

for a finite time horizon, \( t = 1, \ldots, T \), where \( r \geq 0 \) represents the discount factor, subject to System (8) and some terminal conditions \( A(T+1) = A_{T+1} \) and \( D(T+1) = D_{T+1} \). Assuming the cost functions \( E, F, \) and \( G \) are convex the discrete-time maximum principle (see, e.g., Léonard and Long, 1992) yields an implicit condition for the optimal interior treatment policy. (See Appendix A.2 generally and Equation (A13) in particular.) This condition tells us that the change in the
discounted cost of the optimal treatment policies is equal to the number of addicts weighted by their shadow-price (both evaluated along the optimal path). Further results can be derived after specifying the functional forms of the cost functions, as is done in the next section, although it was still not possible to compute the optimal solution trajectory explicitly.

Note that minimizing the sum of convex functions of the state variables and control costs need not entail removing chaotic behavior. Indeed, we put “cost optimal” in quotation marks in part because the objective function does not include a penalty associated with variability in the state. Quotes are also appropriate because the combination of the three convex cost terms (for the dealers, the users, and the control spending) is ad hoc since we have no estimate of their relative importance to the decision maker. The leading coefficient of all three terms is one, which may or may not be realistic. Nevertheless, the control generated by consideration of some such objective function represents an interesting foil for the OGY controls.

4.3 Comparing the Results of Control for Quadratic Costs

The previous sections outlined four treatment control strategies that might be considered, e.g., in conjunction with a reduction in ongoing enforcement intensity that could otherwise induce chaotic behavior. Which is preferred depends on the decision maker’s preferences concerning different types of costs. To give some insight into the relative advantages of these different strategies, Table 1 summarizes for various strategies the NPV of the costs of the state,

\[ \text{CoS} = \sum_{t=0}^{T} (1 + r)^{-t} (E_t(A_t) + F_t(D_t)), \]  

the NPV of the costs of control,

\[ \text{CoC} = \sum_{t=0}^{T} (1 + r)^{-t} G(u_t), \]

and the discounted sum of the squared differences between successive states

\[ \text{MoV} = \sum_{t=0}^{T-1} (1 + r)^{-t+1} \left( (A_{t+1} - A_t)^2 + (D_{t+1} - D_t)^2 \right). \]

which can be interpreted as a measure of the amount of “variability” in the state. We do this for parameter set \( S = \{ \alpha = 0.2977, \beta = 0.4580, a = 0.2000, b = 0.3900, c = 161.1536 \} \) as paradigmatic example of system behavior. We chose an arbitrary initial condition, \((A_0, D_0)\), for numerical simulation. (It should be noticed, how-
ever, that this initial condition has a large influence on some of our measures.) For simplicity we use quadratic cost, i.e.,

\[ E(A_t) = A_t^2, \quad F(D_t) = D_t^2, \quad \text{and} \quad G(u_t) = u_t^2. \] (15)

We choose a time horizon of \( T = 100 \) and a discount factor of \( r = 0.04 \), which is reasonable if the time step is between one quarter and one year.

Note that the theory of dynamical systems postulates that no control is needed if we travel exactly on the stable manifold toward the equilibrium and that no control is needed to remain precisely at the equilibrium forever. In practice, however, we have to maintain the system in a state that is artificially and numerically stabilized (and actually unstable). The associated fluctuations in the equilibrium prevalence create cost for ongoing maintenance for all controlled cases except constant treatment. These are in addition to the cost during the transient phase and are listed separately in Table 1, in part because their magnitude is sensitive to the level of precision used in the computations. The fluctuations also contribute to the state costs and variability, but they constitute a minor proportion of the totals and so are not listed separately.

Table 1 also gives the maximum amount of control used in both the transient and maintenance phases and how long it takes to reach the maintenance phase, defined as the time until the state reaches an \( \varepsilon = 0.0012 \) neighborhood of the equilibrium.

The constant control that is just enough to remove chaos and achieve convergence towards the one-period fixed point \( \hat{X} = (0.0171, 0.0364)^T \) is equal to \( u_c = 0.2023 \). This control quickly (5 periods) pushes the system to a low-state equilibrium8 reducing the cost of the state (i.e., of users and sellers) by 73% and variability (Equation (14)) by 81% relative to the uncontrolled solution, but the cost of control is an order of magnitude greater than with OGY or OGY + steering. Indeed, it is so large that the (equally weighted) sum of state and control costs is actually higher than for the uncontrolled solution. Furthermore, even if the time step is interpreted as one year, \( u_c = 0.2023 \) may be at the outer limits of what treatment can realistically accomplish. E.g., Rydell and Everingham (1994) assumed that each episode of cocaine treatment had a 13% chance of pushing a user out of heavy use and that treating every heavy drug user once per year was about as much treatment as one could do, and even these assumptions have been attacked as potentially too optimistic (Manski et al., 1999).

\(^8\) Note that the time until the system reaches a neighborhood of the fixed point depends on the initial state, i.e. the initial numbers of addicts and dealers.
Table 1: Different measures of system performance for constant control, $u_c$, the OGY control, $u^{OGY}$, a combination of steering and the OGY algorithm, $u^{OGY-ST}$, and costless jump to equilibrium for parameter set $S = \{\alpha = 0.2977, \beta = 0.4580, a = 0.2000, b = 0.3900, c = 161.1536\}$ and for $r = 0.04, \epsilon = 0.0012$ and $T = 100$. Initial conditions are given by $A_0 = 0.13$ and $D_0 = 0.26$.

<table>
<thead>
<tr>
<th></th>
<th>none</th>
<th>$u_c$</th>
<th>$u^{OGY}$</th>
<th>$u^{OGY-ST}$</th>
<th>Costless jump to $\hat{X}$</th>
<th>Costless jump to $\hat{X}_4$</th>
<th>$u^{OGY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{X} \equiv (\dot{A}, \dot{D})$</td>
<td>none</td>
<td>(0.0171, 0.0364)</td>
<td>(0.0557, 0.0706)</td>
<td>(0.0557, 0.0706)</td>
<td>(0.0557, 0.0706)</td>
<td>$X_4$ see text</td>
<td>$\hat{X}_4$ see text</td>
</tr>
<tr>
<td>Cost of transient control</td>
<td>--</td>
<td>0.1895</td>
<td>0.0500</td>
<td>0.0100</td>
<td>--</td>
<td>--</td>
<td>0.1225</td>
</tr>
<tr>
<td>Cost of maintenance control</td>
<td>--</td>
<td>0.8535</td>
<td>0.0002</td>
<td>9.0*10^-6</td>
<td>9.9*10^-6</td>
<td>0.2028</td>
<td>0.0828</td>
</tr>
<tr>
<td>Overall control cost ($CoC$)</td>
<td>--</td>
<td>1.0430</td>
<td>0.0502</td>
<td>0.0100</td>
<td>9.9*10^-6</td>
<td>0.2028</td>
<td>0.2053</td>
</tr>
<tr>
<td>Overall state cost ($CoS$)</td>
<td>0.5556</td>
<td>0.1502</td>
<td>0.3168</td>
<td>0.3017</td>
<td>0.2829</td>
<td>0.2508</td>
<td>0.3006</td>
</tr>
<tr>
<td>Overall sum of control + state costs</td>
<td>0.5556</td>
<td>1.1932</td>
<td>0.3670</td>
<td>0.3117</td>
<td>0.2829</td>
<td>0.4536</td>
<td>0.5059</td>
</tr>
<tr>
<td>Overall variability ($MoV$)</td>
<td>0.1328</td>
<td>0.0251</td>
<td>0.0304</td>
<td>0.0214</td>
<td>0.0414</td>
<td>0.0713</td>
<td>0.0607</td>
</tr>
<tr>
<td>Maximum level of $u_i$ in the transient phase</td>
<td>--</td>
<td>0.2023</td>
<td>0.2149</td>
<td>0.1000</td>
<td>NA</td>
<td>NA</td>
<td>0.2148</td>
</tr>
<tr>
<td>Maximum level of $u_i$ in the maintenance phase</td>
<td>--</td>
<td>0.2023</td>
<td>0.0038</td>
<td>0.0035</td>
<td>0.0036</td>
<td>0.1674</td>
<td>0.1684</td>
</tr>
<tr>
<td>Time until $\epsilon$-neighborhood of $X$ is reached</td>
<td>--</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>21</td>
</tr>
</tbody>
</table>

By contrast, OGY control to the equilibrium $\dot{X} = (0.0557, 0.0706)^T$ has a control cost that is only 5% as large (third column of Table 1). The peak control (0.2149) is actually slightly higher than the constant control, but it is applied for just one period.\[^{[10]}\] The OGY control’s sole aim is to remove chaos. Nonetheless, it also reduces the cost of the state substantially (43%) relative to no control, almost as much as does instantaneously moving to the equilibrium (49%). The sum of the

\[^{[9]}\] The specific performance listed assumes the state jumps first to state (0.05613, 0.0960). If any of the other three states is the first state, the outcomes are similar, with the cost of the state between 0.1731 and 0.1756; variability between 0.0804 and 0.0932; and their sum between 0.2560 and 0.2683.

\[^{[10]}\] It is conceivable that a high outflow rate could be accomplished for a single period even if it could not be done on an ongoing basis if enforcement against users were combined with treatment. In the long run incapacitation is very expensive, but enforcement does have greater capacity than other interventions to focus its impact on a particular point in time (Caulkins, 2000). Still, the need for such a “surge” in control may be a drawback of this OGY solution.
control plus state costs for OGY is less than for no control and much less than for constant control. Not surprisingly OGY also reduces state variability (by 73%), but not quite by as much as the constant treatment strategy does. The reason is simply that it takes longer for the uncontrolled (chaotic) trajectory to accidentally pass through a small tube around the stable manifold and get “hooked”. Once trapped in the small tube around the stable manifold the convergence towards the fixed point is fast.

In this case steered OGY control (column 4) improves on pure OGY in every respect. State costs and variability are reduced slightly, control costs are very low, peak control is only half as intense, and the equilibrium is approached in 2 steps not 7, even faster than with the very aggressive constant treatment control. Steered OGY does not always Pareto dominate pure OGY, but it does approach the equilibrium more quickly and seems to often involve lower total control costs, although the peak control with steered OGY can be large.

Since it was not possible to compute the cost-minimizing trajectory explicitly, two “better than” optimal results are listed as foils in Table 1. Each of these hypothetical (and infeasible) solutions moves the state to its equilibrium in one period and at zero cost. For the first the equilibrium is the one-periodic fixed point \( \hat{X} = (0.0557, 0.0706)^T \). For the second, the equilibrium is the four-period orbit with \( \hat{X}_4 = ((0.0561, 0.0960)^T, (0.0397, 0.0527)^T, (0.0471, 0.0716)^T, (0.0368, 0.0462)^T)^T \). Jumping immediately to the steady state, whether to the 1-period fixed point or the 4-period orbit, reduces the cost of the state more than does either OGY or steered OGY, but only modestly (6-16%) for the cases considered here. By definition the true, feasible optimal control solution could not do any better, so the OGY and steered OGY reductions in the state must be moderately close to optimal.

Ironically the “better than optimal” solutions actually increase variability relative to OGY and OGY + steering. The reason for this surprising result is that our variability measure is the sum of the squared differences in the state, and moving instantly rather than gradually from the initial condition to the steady state gives a very large squared difference in the first period. Moving in two half steps would roughly halve the variability cost since \( 2(0.5(X_0 - \hat{X}))^2 = 0.5(X_0 - \hat{X})^2 \). In effect, these (hypothetical) control strategies move “too quickly”.

It is interesting to observe that the (formerly unstable) four-period orbit gives a lower average state than does the equilibrium \( \hat{X} = (0.0557, 0.0706)^T \). Thus, if the decision maker is willing to tolerate ongoing state variation over time with peaks 1.5 – 2 times as great as the troughs, the decision maker can get 6% fewer
Table 2: Summary of advantages and disadvantages of four types of treatment policies: constant control, $u_c$, the OGY control, $u^{OGY}$, a combination of steering and the OGY algorithm, $u^{OGY+ST}$, and the optimal dynamic control, $u^*$.

<table>
<thead>
<tr>
<th>Type of control</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>no control</strong></td>
<td>No control costs</td>
<td>Chaotic behavior; Cost of state and variability are very high.</td>
</tr>
<tr>
<td>$u_c$</td>
<td>Dynamic policy variation not needed; Takes effect quickly; No variability if control successfully removes chaos; Cost of state is very low.</td>
<td>Eradication of chaos not guaranteed; Both peak and total cost of control are very high.</td>
</tr>
<tr>
<td>$u^{OGY}$</td>
<td>Removes chaos; Variability is moderate; Cost of control and state are low.</td>
<td>Lag – sometimes quite long – before it begins; Requires flexibility of social policy makers; Peak control can be large.</td>
</tr>
<tr>
<td>$u^{OGY+ST}$</td>
<td>Removes chaos; Takes effect quickly; Variability, cost of control, and cost of state are low.</td>
<td>Requires great flexibility of social policy makers; Peak control can be large.</td>
</tr>
<tr>
<td>$u^*$</td>
<td>Cost of states and controls are optimally balanced against each other.</td>
<td>Eradication of chaos not guaranteed; Requires flexibility of social policy makers; Very difficult to compute.</td>
</tr>
</tbody>
</table>

dealers and 19% fewer users, on average, in steady state. This illustrates clearly that the objectives of minimizing the average amount of market activity and minimizing variability in that activity can be in tension. Note: as the last column in Table 1 shows, this 4-period equilibrium can be reached with OGY control, not just a costless jump, although that solution illustrates how pure OGY can sometimes take a long time to reach the equilibrium. Indeed, in this case it takes so long and the equilibrium involves enough ongoing variation that the variability measure is only reduced by about half (54%) relative to the uncontrolled solution.

Table 2 summarizes these observations about the relative advantages of the various control strategies.

5. Conclusions

Street markets for illicit drugs are chaotic in the informal sense of the word, and we have shown that a simple model of the interaction between numbers of users and sellers in those markets can for certain parameter values be chaotic in the formal mathematical sense as well. It is hard to say whether such dynamics are in
fact an important contributor to the unpredictable variation observed in street market activity. An alternative hypothesis is simply that there is an ongoing series of irregular exogenous shocks. Nonetheless it is interesting to consider the possibility that the unpredictable variation may be, at least in part, a product of the markets’ own endogenous dynamics.

The model offers an interesting vehicle for contrasting the effects of different types of control. Static control can help remove chaos, but it can just as easily create it. That is, in some sense neither interventions that remove sellers (enforcement) nor users (primarily treatment, but also enforcement against users) are good or bad per se. Nor can one always say that doing more or less of either type of intervention is the key to reducing chaos. Rather, ideally the control strategy should be sensitive to the particular circumstances of a market and adapt with the current state of that market.

For the parameter set and cost functions examined in greatest depth, we found that so-called OGY control and OGY plus steering seem to be appealing. They reduce the size and variability of the drug markets while expending a modest control effort. A principal objection to pure OGY is that it can take a long time to operate. Steered OGY is superior in that and other respects, but in some cases may require intensities of intervention in the steering phase that are difficult to achieve. Kopel (1997, 1998, 1999) has demonstrated the value of steering + OGY in business optimization models, but to the best of our knowledge this is its first application in a public policy domain.

Methodologically this paper uses global numerical stability analysis methods including the calculation of Lyapunov spectra and the fractal dimension (Khibinik et al., 1993; Medio, 1992; or Schaffer et al., 1988). Additionally intensive local bifurcation analysis was performed and illustrated. More generally, it uses both the theory of nonlinear dynamics and chaos and points to the potential to merge steering, OGY and optimal control by developing objective functions that includes costs for both the state and variation in the state.

Substantively one might imagine other drug market models that exhibit interesting chaotic behavior. For example, drug epidemics seem to strike in unpredictable ways. (Washington D.C. had a sharp PCP epidemic in the mid-1990s; nearby Baltimore did not.) Likewise there is strange variation from year to year in production capacity in source countries that could conceivably stem from lagged and myopic response to price fluctuations generated by downstream enforcement operations.
Appendix

A.1 Deriving the Conditions for the OGY Control

In vector form System (8) can be rewritten as

\[
X_{t+1} = \Psi(X_t) = \begin{pmatrix} 1 - \alpha & 0 \\ 0 & 1 - \beta \end{pmatrix} X_t + \begin{pmatrix} a \\ \frac{a}{b} \end{pmatrix} f(X_t),
\]

(A1)

with \(0 < a < b\); \(c \in (c_1, c_2)\) and \(0 \leq \alpha, \beta \leq 1\),

where

\[
X_t = \begin{pmatrix} A_t \\ D_t \end{pmatrix},
\]

(A2)

and where \(f\) is defined by Equation (2). The one-periodic fixed point evaluated for a particular \(\alpha\) is given by

\[
\hat{X}^\alpha = \begin{pmatrix} \hat{A} \\ \hat{D} \end{pmatrix} = x \begin{pmatrix} 1 \\ \frac{a\beta - a\beta}{a\beta} \end{pmatrix},
\]

(A3)

where \(x\) is the solution of the equation

\[
\alpha x + \alpha x e \begin{pmatrix} a\beta - a\beta \\ a\beta \end{pmatrix} - a = 0.
\]

(A4)

The fixed point (A3) must have the properties of a saddle point and, thus, exhibit one stable direction. Let \(\lambda_s\) and \(\lambda_u\) be the stable and unstable eigenvalues (|\(\lambda_u| > 1 > |\lambda_s|\)) of the map \(\Psi\) at the desired fixed point \(\hat{X}^\alpha\). Furthermore, let \(e_s\) and \(e_u\) be the unit vectors in the stable and unstable directions of the fixed point, respectively. Assuming that the eigenvectors, \(e_s\) and \(e_u\), are normalized, i.e. \(|e_s| = |e_u| = 1\), one can find a pair of vectors \(y_u\) and \(y_s\)—called the contravariant basis vectors—that are perpendicular to the unstable and stable axis, respectively (see Ott et al., 1990; Shinbrot, 1993). Near the fixed point \(\hat{X}^\alpha\) we can use following linear approximation for the map \(\Psi(X_t)\)

\[
X_{t+1} = \Psi(X_t) \approx \hat{X}^\alpha + \hat{J}(X_t - \hat{X}^\alpha),
\]

(A5)

where \(\hat{J}\) denotes the Jacobian matrix of the function \(\Psi\) evaluated at \(\hat{X}^\alpha\). Then we change the control parameter from its original value \(\alpha\) to some other value \(\alpha \leq \tilde{\alpha} \leq 1\).

The fixed point coordinates will shift to some nearby point \(\hat{X}^{\tilde{\alpha}}\) which will be approximated for small values of \(\tilde{\alpha}\) by

\[
\tilde{\alpha} = \frac{\partial \Psi}{\partial \alpha}|\hat{X}^\alpha \equiv \frac{\hat{X}^{\tilde{\alpha}}}{\tilde{\alpha}}.
\]

(A6)
Assume that a $t$ exists such that $X_t$ falls near the desired fixed point $\hat{X}^\alpha$ so that (A6) applies. We then attempt to pick a parameter $\tilde{\alpha}_t := \alpha + u_t^{\text{opt}}$ such that $X_{t+1}$ falls on the stable manifold of $\hat{X}^\alpha$. That is, we choose $\tilde{\alpha}_t$ such that $y_u \circ X_{t+1} = 0$ holds. This yields the following rule to compute $\tilde{\alpha}_t$:

$$\tilde{\alpha}_t = \alpha - \frac{\lambda_u y_u \circ (X_t - \hat{X}^\alpha)}{y_u \circ \omega} \tag{A7}$$

which we use when $\alpha \leq \tilde{\alpha}_t \leq 1$. Otherwise, we set $\tilde{\alpha}_t = \alpha$. As soon as $X_{t+1}$ exactly falls on the stable manifold of $\hat{X}^\alpha$ we can set the parameter perturbations to zero, and, for subsequent time, the orbit will approach the fixed point at the geometrical rate $\lambda_s$.

### A.2 Deriving the Conditions for the Optimal Control

For performing a single decision-maker’s optimization by minimizing the additive separable cost function (14) subject to system (8) (applying the discrete-time minimum principle as outlined by e.g. Léonard and Long (1992, pp.130–131)), we have to set up the Hamiltonian function, $\mathcal{H}$, for any time period $t \in \{1, \ldots, T\}$,

$$\mathcal{H}_t = (1 + r)^{-t} \left( E(A_t) + F(D_t) + G(u_t) + \pi_u \left((1 - \alpha - u_t)A_t - A_{t+1} + af_t\right) + \pi_{1t} \left((1 - \beta)D_t - D_{t+1} + b \tilde{f}_t\right) \right) \tag{A8}$$

where $f = f(X_t)$ is defined by Equation (2), $X_t$ by Equation (A2), and where the terms $\pi_u$ ($i = 1, 2$) denote the costates. The admissible set of controls is defined by $U = \{u | u \geq 0, 1-\alpha-u \geq 0\}$. Hence, we have to set up the Lagrangean function,

$$L = \mathcal{H}_t + \lambda_{1t} u_t + \lambda_{2t} \left(1 - \alpha - u_t\right), \tag{A9}$$

where the nonnegative Lagrangean multipliers, $\lambda_i$ ($i = 1, 2$), are piecewise continuous on $1 \leq t \leq T$. Thus, the following complementary slackness conditions hold:

$$\lambda_{1t} \geq 0, \quad u_t \geq 0 \quad \text{and} \quad \lambda_{1t} u_t = 0, \tag{A10}$$
$$\lambda_{2t} \geq 0, \quad 1-\alpha-u_t \geq 0 \quad \text{and} \quad \lambda_{2t} \left(1 - \alpha - u_t\right) = 0. \tag{A11}$$

Applying the total differential to the Lagrangean (A9) yields the Lagrangean minimizing condition for all time periods $t \in \{1, \ldots, T\}$ and for all interior controls,

$$\frac{\partial L}{\partial u_t} = 0 \iff (1 + r)^{-t} G'(u_t) - \pi_{1t} A_t = 0, \tag{A12}$$

and the equations for the adjoint variables,
\[
\frac{\partial \mathbf{L}}{\partial A_{t+1}} + \frac{\partial \mathbf{L}_{s+1}}{\partial A_{t+1}} = 0 \quad \text{and} \quad \frac{\partial \mathbf{L}}{\partial D_{t+1}} + \frac{\partial \mathbf{L}_{s+1}}{\partial D_{t+1}} = 0. \tag{A13}
\]

Since \( f_A = A \phi \) and \( f_D = -D \phi \), we can rewrite Equations (A13) as follows

\[
\pi_1 = (1 + r)^{-t} E'(A_{t+1}) + \pi_1(t+1) \left(1 - \alpha - u_{t+1} + a A_{t+1} \Phi_{t+1}\right) + \pi_2 A_{t+1} A_{t+1}, \tag{A16}
\]

\[
\pi_2 = (1 + r)^{-t} F'(D_{t+1}) - \pi_1(t+1) a D_{t+1} \Phi_{t+1} + \pi_2(t+1) \left(1 - \beta - b D_{t+1} \Phi_{t+1}\right). \tag{A17}
\]

As long as we can assure that there exists no quadruple \((A_t, D_t, \pi_{1_t}, \pi_{2_t})\) satisfying the conditions for the existence of a solution at the border of the admissible set, \( U \),

\[
\mathbf{H}_f(u^* = 0) < 0, \tag{A18}
\]

\[
\mathbf{H}_f(u^* = 1 - \alpha) < 0, \tag{A19}
\]

the optimal dynamic control falls into the interior of the admissible set, \( U \). Together with System (8) the costate equations, (A16) and (A17), and the Hamiltonian maximizing condition, (A13), constitute the associated canonical system.
References


22


