The Dynamic Process of Dynamic Modeling: 
The Cocaine Epidemic in the United States

D.A. BEHRENS and G. TRAGLER

Working Paper No.248
November 2000
The Dynamic Process of Dynamic Modeling: The Cocaine Epidemic in the United States

Doris A. Behrens
Department of Operations Research and Systems Theory
Vienna University of Technology
and
Department of Economics, University of Klagenfurt

Gernot Tragler
Department of Operations Research and Systems Theory
Vienna University of Technology

Abstract: This paper reviews several recent dynamic models of the current U.S. cocaine epidemic (both uncontrolled and optimally controlled) which differentiate between two levels of use (“light” and “heavy”). Even though all these models have their origin in a study carried out at the RAND corporation’s Drug Policy Research Center (DPRC) in the early nineties, each has been developed by extending or refining another. Apart from pointing to interesting policy conclusions derived from the analysis of these models, this paper also demonstrates the fact that the development of dynamic models of illicit drug consumption is itself a dynamic process where subsequent refinements lead to subsequently increasing quality and reliability of the resulting policy conclusions.

Keywords: Illicit Drug Use, Demand Control, Nonlinear Dynamic Systems, Hopf-Bifurcations, Optimal Control.

* This research was partly financed by the Austrian Science Foundation (FWF) under contract No. P11711–OEK (“Dynamic Law Enforcement”).

• mailing address: Argentinierstrasse 8/119, A-1040 Vienna, Austria; dbehrens@eos.tuwien.ac.at
1. Introduction

Illicit drug use and related crime have imposed significant costs on the U.S. and various source and transshipment countries for a number of years. A variety of control strategies exist including prevention, treatment, and various forms of enforcement, so a fundamental question in drug policy is how should scarce resources be allocated between these programs. Analysts have sought to inform this decision by estimating the cost-effectiveness of different interventions. The greater part of this work has made estimates only for a particular point in time, concluding, for example, that in 1992 domestic enforcement was three times more cost-effective than border interdiction (Rydell et al., 1996). Earlier studies that used dynamic models have not focused on cost-effectiveness (see, e.g., Schlenger, 1973; Levin et al., 1975; Gardiner and Shreckengost, 1987; and Homer, 1993a,b).

Behrens et al. (1999) contribute to the effort of understanding drug use and how it responds to drug control interventions by introducing a simple continuous time model of drug demand that incorporates a feedback effect of the current prevalence, or level, of use on initiation in new use. Analyzing the model generates important new insights into how epidemics of drug use should be studied and, to the extent that such a simple model can be trusted, how they should be controlled. This model has received refinements along different lines.

First, in Behrens et al.’s original model, the deterrent effect on initiation was governed by the current number of heavy users. This approach kept the model’s complexity low enough to simplify the analysis, while featuring the rather unrealistic assumption that when a heavy user exited the user population, his or her contribution to deterring initiation went immediately to zero. In other words, there was absolutely no memory of past bad experiences with heavy use even though in reality not all knowledge of a heavy user’s bad experiences disappears the moment the individual exits the population, particularly if the exit is by physical death from drug use. More recent models avoid this problem by introducing a third state (in addition to the numbers of light and heavy users) which reflects some sort of memory of drug abuse.

In this paper we review Behrens et al.’s (1999) model as well as the family of models stemming from its refinements (Behrens et al., 2000a, 2002). Apart from the description of these dynamic models, we will point to the major policy conclusions resulting from their analyses. Hence, the purpose of this paper is twofold. In an application-oriented vein, we want to give concrete examples of dynamic models that have been parameterized with data of the current U.S. cocaine epidemic. In a more philosophical vein, we point to the fact that developing dynamic models is itself a highly dynamical process due to the fact that data sources are permanently gaining higher quality, that the understanding of the underlying epidemic processes
grows, and that the recent developments of the tools needed for the analysis of dynamic models (both hard- and software) lead to subsequent improvements of the models (and hence to a higher quality / reliability of the resulting policy conclusions).

2. A Dynamic Model of the U.S. Cocaine Epidemic
   Including an Endogenous Feedback on Initiation

Since there is enormous heterogeneity across drug users with respect to the rates of consumption and since the average rate of consumption for a population can change over time, tracking trends in total consumption (which is a reliable measure for the size of a drug problem according to Rydell et al., 1996) requires separate modeling of the numbers of users at different levels, or intensities, of drug use. Ideally, one would model the whole spectrum of consumption behavior, from occasional use in small amounts up to frequent use in large amounts, but data limitations make that infeasible.

Everingham and Rydell (1994) recognized this tension and suggested that, at least for cocaine, a simple dichotomous distinction between “light” and “heavy” users is useful. They operationalized the distinction using data from the National Household Survey of Drug Abuse (National Institute on Drug Abuse, 1991) which measures the prevalence of cocaine use among the U.S. household population. In particular, people who reported using cocaine “at least weekly” were defined to be heavy users, while those who consumed at least once within the last year but used less than weekly were called “light users”. The average heavy user consumes cocaine at a rate approximately seven times that of an average light user and exhibits substantially greater adverse consequences associated with that drug use.

A significant limitation of the Everingham and Rydell (1994) model was that initiation was scripted. Future projections and policy simulation exercises were predicated on a fixed projection of future initiation that is insensitive to the course of the drug epidemic. That is problematic because the current prevalence of use significantly influences initiation rates. In particular, most people who start using drugs do so through contact with a friend or sibling who is already using. Indeed, the metaphor of a drug “epidemic” is commonly used precisely because of this tendency for current users to “recruit” new users. If that were the only mechanism by which current use affected initiation one might expect initiation to increase monotonically. Musto (1987) has argued that, in addition, knowledge of the possible adverse effects of drug use acts as a deterrent or brake on initiation. He
has hypothesized that drug epidemics eventually burn out when a new generation of potential users becomes aware of the dangers of drug abuse and, as a result, does not start to use drugs. Whereas many light users work, uphold family responsibilities, and generally do not manifest obvious adverse effects of drug use, a significant fraction of heavy users are visible reminders of the dangers of using addictive substances. Hence, one might expect large numbers of heavy users to suppress rates of initiation into drug use. It seems plausible that any reasonable model of an endogenous initiation might have the following properties.

1) The rate at which current users “recruit” initiates is proportional to the number of light users. It is assumed that heavy users do not recruit initiates because they manifest ill effects of drug use and/or because they have been using long enough, so that they are older and socially distant from youth in the prime initiation ages.

2) The rate at which current light users “recruit” initiates is moderated by the “reputation” or image the drug has, and the reputation is governed by the relative number of heavy and light users, not the absolute number of heavy users. Even if there were a number of heavy users, the drug might appear benign if they were buried in a sea of (relatively happy) light users.

3) Although most new users are “recruited”, for others the impetus to use is internal. In the jargon of diffusion models (cf. Gould, 1970), these individuals are “innovators” who initiate on their own for the sake of curiosity, by shifting from other drugs, or for some other reason, but not through the urging of someone who is already a user.

About 60 functional forms catching these features have been investigated, where Behrens et al. (1999) choose one of those 5 functional forms which give the best system performance with respect to minimization of the squared differences between modeled and observed initiation data from 1970–1991.

The rest of the so-called LH model (see Behrens et al., 1999) is essentially a continuous time analogue of Everingham and Rydell’s (1994) model. In this model, the population is divided into three groups: non-users, light users and heavy users (cf. the beginning of this section). The number of non-users is assumed to be large enough compared to the number of users to behave like a constant and does not need to be modeled explicitly (see Everingham et al., 1995). The flow rates from one state to another are assumed to be proportional to the source states and are computed as the time-continuous equivalents of the Everingham-Rydell estimates (Everingham and Rydell, 1994, p. 43).
In addition to the endogenously modeled initiation, another difference between Behrens et al.'s (1999) model (illustrated in figure 1) and that of Everingham and Rydell (1994) is that in the latter, the outflow from heavy use is divided into a flow out of use altogether (currently denoted by the “rate of desistance”) and a flow back into light use. Behrens et al. (1999) dropped the latter flow for both theoretical and practical considerations. Theoretically, a flow from heavy to light use coupled with the Markov assumption implies that former heavy users who have de-escalated to light use and light users who had never been heavy users are indistinguishable. But probably it is easier to relapse into heavy use than to enter the state for the first time. Hence, Behrens et al. (1999) prefer to have only a flow from heavy use to non-use and view that rate as net of relapse.

Before one can draw policy conclusions from such a dynamic model of illicit drug use, one has to make sure that the observed drug epidemic can be replicated by that model. (In the terminology of a mathematician, this is a “necessary” but not “sufficient” condition for further analyses.) As figure 2 shows, the fit of the modeled epidemic is not perfect; the historical data reflect a higher, sharper peak in light use. Nevertheless, the similarity is striking given that the actual epidemic was subject to a varying set of drug control interventions over time that could be responsible for deviations from the model’s uncontrolled path. Likewise, idiosyncratic historical events, such as Len Bias’ death and the sharp increases in prices in late 1989, could account for some of the differences between historical and modeled data. And, of course, we cannot expect a perfect fit for a relatively simple model of a very complicated process such as the current U.S. cocaine epidemic.

In their extensive analyses, Behrens et al. (1999) found that omitting the feedback effects of prevalence on initiation was of relatively little consequence for the analysis of the effectiveness of treatment and enforcement at a particular
point in time, as Rydell and Everingham (1994) did. It is of enormous consequence, however, for understanding how effective prevention programs are or for understanding how the effectiveness of an intervention such as treatment might vary over the course of an epidemic. Although Behrens et al. (1999) did not investigate an optimal control model but a purely descriptive model, they did derive a number of interesting results with respect to the nature of drug control interventions by means of simple sensitivity analysis. For example, they found that different strategies are most effective at different stages of an epidemic, and one would expect the optimal mix of interventions to depend significantly on the course and status of the epidemic. More precisely, Behrens et al. (1999) hypothesized that prevention programs might be most effective at early stages of the epidemic, when most users are light users, while treatment programs may be most effective when a greater fraction of users are heavy users, as is typical for later stages of an epidemic.

Inasmuch as it makes sense to vary the mix of interventions over the course of an epidemic, the authors intended to apply optimal control theory to their modeled epidemic – as a topic of further research in this area. Even though this extension seems to be the “next logical step”, it should be noticed that, generally, the derivation of the optimal choice of one or several controls in a dynamic model is a very complex and sophisticated task – also for models like the one introduced by Behrens et al. (1999), which in other respects are quite “simple”. The following section deals with this optimal control “enterprise”.

Figure 2: Time paths of the continuously modeled U.S. cocaine epidemic and the smoothed historical data (Everingham and Rydell, 1994)
3. An Optimal Control Model

From Behrens et al.’s (1999) descriptive model we conclude that drug control interventions should change over time – especially over the course of a drug epidemic. The way they do so crucially depends on the choice of the interventions, on the objective, and finally on the choice of restrictions on the drug control budget.

Behrens et al. (2000a) formulated and solved an optimal control model to derive optimal intertemporal treatment and prevention spending decisions under three different assumptions with respect to restrictions on the drug control budget. In particular, the original model (presented in the previous section and illustrated in figure 1) has been extended in that two of the flows are influenced (“controlled”) by suitable control instruments: (primary) prevention decreases initiation by a certain percentage, while treatment of heavy users increases their rate of desistance as illustrated in figure 3. The objective chosen by Behrens et al. (2000a) was to minimize the total social costs rather than to maximize social welfare in the sense that drug users’ consumer surplus was excluded from the objective functional. The total social costs included both the social costs caused by illicit drug use and the additional monetary costs of the control measures (i.e., treatment and prevention spending). An alternative objective for the LH model has recently been presented by Kaya (2000), namely to reach some predetermined target in optimal time.1

As already mentioned above, Behrens et al. (2000a) considered three different assumptions for restricting the drug control budget. These three cases can be described as follows:

(C1) the budget is constrained to be proportional to the size of the cocaine problem,2 and the proportions of that budget going towards treatment and prevention, respectively, are chosen once and fixed for all time;

(C2) the budget is chosen as in case (C1), but its allocation between treatment and prevention can be varied over time; and

(C3) the budget is unconstrained in that both treatment and prevention spending can be chosen to be any non-negative number at all times. Theoretically, this case is the most reasonable one, because at some stages of the epidemic, high expenditures may be useful, while at other stages spending less money may be preferable. Practically, however, the implementation of

---

1 In particular, Kaya (2000) considered the problem of finding a time-optimal control to get from some initial state (i.e., initial numbers of light and heavy users) to a target state.

2 According to Rydell et al. (1996), total consumption is a reliable measure for the size of a drug problem.
the optimal solution to this problem may cause problems because the optimal expenditures can be considerably high or vary significantly over time. It should also be noted that this model is more appropriate than one with a constrained budget if treatment and prevention resources are not allocated from a single pot.

Comparing the results of these three constrained and unconstrained optimization models sheds light on how different forms of political constraints affect drug control. Insights of the optimally controlled LH model (Behrens et al., 2000a) include:

1. Applying static interventions to a dynamic process may be counter-productive. That means, control measures, such as treatment and prevention, are most appropriate for specific stages of a drug epidemic, and budget allocations across these measures should change over time. For instance, prevention does best when there are relatively few heavy users, i.e. in the beginning of an epidemic. Treatment, on the other hand, is relatively more efficient at supporting the decline of drug abuse later in the epidemic (see figure 4).

---

3 Tragler et al. (2001) derived the optimal solution for an alternative optimal control model of the U.S. cocaine epidemic. Their model is different in that they consider “average” users (i.e., there is no distinction between different levels of use), and the controls are treatment and price raising enforcement. Tragler et al. (2001) show that, if initiation into drug use is an increasing function of the current number of users and control measures are implemented early in the drug epidemic, then it is optimal to use very large amounts of both enforcement and treatment to eradicate the epidemic. In other words, one would need a very large budget to pursue an optimal control of a drug epidemic which is still in its early stages, i.e. small. In such a case, the per-user budget is enormous, and one can imagine that the optimal policy will not be implemented because the public would not accept very large expenditures for a problem which is hardly visible.
2. The transition period, when it is optimal to use extensively both prevention and treatment, is brief (figure 4).

3. Some control, even a "dumb" one such as is embodied in the case in which the budget is not only constrained in total size, but also the proportions of that budget going towards treatment and prevention are chosen once and fixed for all time, does better than no control (see figure 5).

4. People who perceive drug use to be costly for society should favor greater drug control spending per gram consumed and allocate a greater proportion of that spending to prevention. Generally, it would be most effective to provide very large financial resources for control measures right from the onset of an epidemic (for prevention programs), even if it might be difficult to justify this behavior by the "magnitude of the problem" at the time.

5. Total social costs increase dramatically if control is delayed (see figure 6).

Looking at these insights from the Behrens et al. (2000a) study more carefully, we see that most of the results are as one would have expected them to be (e.g., controlling the epidemic is good, delaying the control is bad, the controls should vary over the course of an epidemic, etc.). On the other hand, the conclusions that

---

**Figure 4:** Prevention and treatment spending in billion dollars for the unrestricted optimal control model (described in case (C3) above)
we draw from figure 4 are to some extent awkward: it follows that it is optimal to stop prevention at a time, when the epidemic is still in its early stages, while treatment should not be implemented before the epidemic has “matured” somewhat. This result follows from the implicit model assumption that large numbers of heavy users are not only bad in the sense that they consume at high rates and hence impose large costs on society, but also good in the sense that they tend to discourage initiation. In other words, heavy users do impose costs in the near term, but they also generate a perverse sort of “benefit” for the future by reducing current initiation and thus future use. Since the timing of treatment and the reputation of the drug strongly interact in the framework of the model described above, the reputation mechanism deserves re-consideration. One possible refinement of the LH model can be obtained by assuming that the reputation influencing initiation is not a function of the current number of heavy users but rather of the memory of past heavy users; this framework is the content of the next section.

Figure 5: Total quantity consumed during the current U.S. cocaine epidemic as well as controlled quantities for the different budget rules.
4. Modeling a Memory of Drug Abuse

The extension of the reputation function to be a function of the decaying memory of heavy users, carried out by Behrens et al. (2002), has obvious appeal. As we mentioned already earlier in this paper, all knowledge of the negative experiences of a heavy user is unlikely to disappear the moment the individual exits the heavy user population, particularly if the exit is by physical death from drug use (as opposed to ceasing use or moving out of the area). In the LH models described in sections 2 and 3, removing a heavy user immediately erased all memory of that individual, so it sometimes appeared preferable to allow a person to suffer rather than to help that person to recover. In other words, the benefits of helping heavy users directly were outweighed by the cost of not being able to set an example with the help of their suffering. Such inhuman policies are most likely to disappear in a framework where past users can be remembered.

From a technical point of view, this refinement of the LH model requires a third state, E (the number of so-called “ever-heavy users”), so the analysis becomes more complicated (see figure 7 for an illustration of this so-called LHE model). It is worth noting that this model extension with an additional state does not significantly improve the system performance in the descriptive (i.e., uncontrolled) case. That means, the time paths of the numbers of light and heavy users look pretty much like those for the original LH model (see figure 2).
Initiation rate of escalation rate of desistance altered by treatment

Figure 7: Flow diagram of the optimally controlled LHE Model

However, in the optimally controlled LHE model, the memory removes the seemingly perverse results of the LH model that the presence of a heavy user can be so valuable as a deterrent that successfully treating such users actually increases consumption in the long run. In particular, treatment is no longer counter-productive unless one fails to keep up the memory of the adverse consequences of abuse in a sufficient way (see Behrens et al., 2002).

An interesting question investigated in the Behrens et al. (2002) paper is the fascinating interaction between a society’s present-orientation, its ability to remember the past, and the occurrence of cycles in the future. Behrens et al. (2002) re-discover the old adage that “those who forget the past are condemned to repeat it”. More precisely, the greater the deterrent power of memories of drug abuse, the less likely society is to wind up with a chain of drug epidemics. Additionally, they verify that it can be desirable to relive past epidemics – at least for myopic decision-makers. Or, to put it in simple terms, “for those who forget the past and over-value the present, it may be optimal to have their future recreate the past”. Finally, it is shown that it is optimal to apply prevention throughout the

---

4 Note that cycles (i.e., repeated drug epidemics) may also occur in the LH model presented in section 2; for details, see Behrens et al. (1999).
epidemic because moderating the contagious aspect of initiation reduces the likelihood of cycles and instability.

The results derived from the LHE model and their comparison with the conclusions of the LH study suggest that the extension to include a memory of people who have been ever heavy users significantly improves the model performance. Nonetheless, the LHE model is not the ultimate model for including the implementation of a memory of drug abuse. The following section deals with another model refinement.

5. Refining the Modeling of the Memory of Heavy Use

The major drawback of the LHE approach is that three individuals who use heavily for one day, one year, and one decade, say, would all contribute the same amount to the memory of heavy use. In reality, however, the longer an individual is addicted, the more problems he or she experiences, the greater the costs imposed on others, the more people there are who witness the behavior, etc. So an appealing alternative is to base the negative reputation of an addictive substance not on the memory of the number of people who ever used heavily, but instead on the memory of the number of heavy user years, i.e. the number of person years spent in heavy use.

Behrens et al. (2000b) have started to investigate such a model, where the number of ever-heavy users (E) is replaced by the number of heavy user years (Y). This model, which is referred to as the LHY model, is illustrated in figure 8. From the preceding discussion it is clear that the LHE and the LHY models differ, and that the LHY model provides a more realistic model formulation. A comparison of these two models' flow diagrams (figures 7 and 8), however, does not reveal the fact that the analysis of the LHY model is also more difficult.5

What both models have in common, however, is the difficulty of estimating the parameters pertaining to the states E and Y, respectively, for which there are no tangible quantities.6 Nevertheless, at least for the LHY model, this problem has been resolved recently in a parameter estimation study with data on the current U.S. cocaine epidemic (Knoll and Zuba, 2000).

5 In contrast to the LHY model, the LHE model breaks into two parts (the LE and the H part) and, hence, the relevant dynamics for the descriptive case (but NOT for the optimally controlled case!) can be transformed into a two-dimensional dynamical system, i.e. the model may be investigated in the plane allowing a more thorough analysis, e.g., with respect to cycles. This simplifies the analysis of the LHE model significantly.

6 Obviously, this problem does not arise in the original LH models.
In Behrens et al. (2000b), we find a thorough stability analysis of the uncontrolled LHY model’s dynamics. The results obtained so far suggest – among other things – that drug prevention can temper drug prevalence and consumption, and can avoid the re-occurrence of a drug epidemic. Furthermore these results evidence the correlation between the deterrent power of negative experiences with drug abuse and the rate of forgetting them by providing a functional form for this phenomenon. The insights derived so far are general enough to allow a detailed characterization of what types of drugs – in terms of the probability of escalating to heavy use and the length of a typical addiction career – are most prone to generate cyclic, i.e. re-occurring, drug epidemics. More than that, the LHY model even allows fairly general statements on epidemics of delinquent behavior with a feedback effect of prevalence on initiation.

6. Summary and Conclusions
Four dynamic models of the current U.S. cocaine epidemic have been reviewed in this paper. All differentiate between two levels of use (“light” and “heavy”) and are based on Everingham and Rydell’s (1994) model, but each has been developed
by extending or refining another. Section 2 presents the so-called LH model by Behrens et al. (1999) which extends Everingham and Rydell’s (1994) model by introducing an endogenous function where light and heavy users feed back on initiation into light use (“infection” by light users versus “deterrence” by heavy users). As is demonstrated in Figure 2, even such a “simple” model as the LH model is to some extent capable of reproducing such a “complicated” process as the current cocaine epidemic in the U.S.

The optimally controlled LH model by Behrens et al. (2000a) has been reviewed in Section 3. The results derived from that study are interesting, and most of them are not extremely surprising (e.g., controlling the epidemic is good, delaying the control is bad, the controls should vary over the course of an epidemic, etc.). Still, one of the results (Figure 4) suggested that this model would need some improvement: There are times in an epidemic where we wish a person to suffer more that to help him (her) since the physical existence of a heavy user significantly diminishes initiation. Behrens et al. (2000a) recognized that these seemingly “perverse” results were caused by the specific choice of the feedback function in the LH model, in particular, the assumption that heavy users would only contribute to a bad reputation of the drug as long as they are heavy users.

For the price of introducing a third state (the so-called “ever-heavy” users), Behrens et al. (2002) were able to formulate a refined model (described in Section 4). One of the interesting questions investigated for this model is the fascinating interaction between a society’s present-orientation, its ability to remember the past, and the occurrence of cycles in the future.

In the LHE model for the (negative) reputation of the drug it did not matter, how long a heavy user was using heavily. This problem has been resolved by replacing the number of ever-heavy users with the number of heavy user years. Section 5 gave a short introduction into this LHY framework (Behrens et al., 2000b).

We conclude with three brief remarks. First, models such as the LHY model may be general enough to be applied to other drugs (both licit and illicit) and even to other forms of delinquent behavior (if they include a feedback effect of prevalence on initiation).

Second, it is obvious from the studies reviewed in this paper that dynamic models provide an important contribution to the problem of designing better drug control policies, because drug epidemics are, by definition, dynamic.

Third, as follows from this paper, the development of dynamic models of illicit drug consumption is itself a dynamic process, in which existing models are extended or refined. This improvement, however, crucially hinges upon an improvement of the data sources, a better understanding of the underlying epidemic.
processes, or new developments of the tools needed for the analysis of dynamic models (both hard- and software).

References


