Controlling the US Cocaine Epidemic: Prevention from Light vs. Treatment of Heavy Use[+]

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

Jonathan P. Caulkins [°]
Carnegie Mellon University
H. John Heinz III School of Public Policy and Management
and
RAND Drug Policy Research Center

Gernot Tragler and Gustav Feichtinger[*]
Vienna University of Technology
Department of Operations Research and Systems Theory

Abstract:

Since drug use and related problems substantially change over time, it seem plausible that drug interventions should vary too. Static interventions applied to a dynamic process may be counter-productive. Therefore we formulate the choice between treatment and prevention spending in the framework of dynamic optimal control. Prevention is the most appropriate control policy when there are relatively few heavy users, i.e. in the beginning of an epidemic. Treatment, however, is most sufficient to support the decline of drug abuse optimally. Additionally these models are able to generate a number of interesting insights, among which are (1) costs of interventions as well as social cost associated with the quantity consumed increase with the delays in the starting year of control, (2) people who perceive drug use to be costly for society should favor greater drug control spending per gram consumed and allocate a grater proportion of that spending to prevention.

[+] This research was partly financed by the Austrian Science Foundation under contract No. P11711-SOZ („Dynamics of Law Enforcement“), the National Consortium on Violence Research, and by the US National Science Foundation under Grant No. SBR-9357936. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation

[°] 5000 Forbes Ave., Pittsburgh, PA 15213; email: caulkins+@andrew.cmu.edu

[*] corresponding author: Argentinierstraße 8/119, A-1040 Vienna, Austria;
email: or@e119ws1.tuwien.ac.at

Contents

Chapter 1

Introduction

Illicit drug use and related crime have imposed significant costs on the US and various source and transshipment countries for a number of years. More recently, drug problems have grown in other industrialized countries to the point that, in Stares' (1996) terms, drugs have become a „global habit." Hence it is important to understand drug use and how it responds to drug control interventions, including prevention, treatment, and various forms of enforcement, among others. Drug use and associated problems change substantially over time. There are waves of grater and lesser drug consumption, and the use of the metaphor of a drug „epidemic" points out the dynamic process. Therefore it seems plausible that there is no single, static strategy solving the drug problem and that the optimal mix of interventions should vary over time.

Goal of this paper is to answer how the allocation of resources to treatment and prevention should vary over the course of a drug epidemic. We do this by applying the tools of optimal control theory to a continuous time dynamic model of drug demand that incorporates a feedback effect of the current prevalence or level of use on initiation into new use. This paper will not draw ultimate conclusions about the topic of dynamic drug control. Drug prices are assumed to be constant and enforcement is exogenously determined and treated outside the models presented here. Hence, this paper should be viewed as a first approach to an important topic area, and it should be considered together with parallel efforts (e.g. Tragler, et al.; 1997) and subsequent research. Nevertheless, the models produce important and interesting insights for how policy should be pursued and evolve over time.

The models are parameterized with data from the current US cocaine epidemic because of its magnitude and because data on that epidemic are relatively good, but it is at least plausible that the qualitative conclusions generalize to similar drugs in similar contexts. Occasionally we make reference to specific numerical findings where appropriate, but those should be understood as merely suggestive.

Chapter 2

The Current Cocaine Epidemic

Cocaine use started growing in the US in the late 1960s, the number of users peaked in the early 1980s, but total consumption remained near its mid-1980s peak for more than a decade. Thus trends in total consumption did not and do not mirror trends in overall prevalence. This raises the question of what summary measure best reflects trends in the magnitude of a drug problem.

Though prevalence (the number of cocaine users) is often used as a measure for the size of the drug problem, it is not particularly a good one (Caulkins and Reuter, 1997). There are many others, e.g. the total amount of consumed cocaine closely related with drug spending, the number of people requiring treatment, drug-related crime and the number of drug-arrests, the social cost of cocaine, and the number of drug-related emergency room episodes. Rydell et al. (1996) argue, though, that the quantity or weight consumed has advantages as a general purpose scalar measure of the size of the problem.

But modeling the whole spectrum of consumption behavior, from occasional use in small amounts up to frequent use in large amounts, is infeasible due to data limitations. Everingham and Rydell (1994) recognized this tension and suggested that, at least for cocaine, a simple dichotomous distinction between „light“ and „heavy“ use is sufficient. Based upon the *National Household Sur-*

vey of Drug Abuse, *NHSDA*, which measures the prevalence of cocaine use among US household population, the frequency of cocaine consumption delivers a selective criterion to distinguish between „light“ and „heavy“ use. People who declared to have consumed cocaine „at least weekly“ are defined to be heavy users, those who consumed at least once within the last year, but less than weekly are stated to be light users. Speaking in terms of consumption the average heavy user consumes cocaine at a rate approximately seven times that of the average light user. The upward trend in consumption by heavy users, which opposes much higher costs to society, outweighs the downward trend in consumption by light users.

Susan S. Everingham and C. Peter Rydell set up a Markovian model of population flows into and out of light and heavy use where initiation was scripted and prevalence closely related to consumption. Based upon data from the *NHSDA* (1991) they selected the transition parameters that determine the flows between those states to match the historical data. They used this model of cocaine-demand to understand what has happened to date in the current cocaine epidemic, to project the future under different incidence scenarios and to compare effectiveness of various drug control policies (Everingham and Rydell, 1994; Rydell and Everingham, 1994).

In order to observe the dynamic flows into and out of the states of light and heavy use incorporating a different incidence scenario, and to investigate the behavioral content we set up a time-continuos version of the first-order difference equation model introduced by Everingham and Rydell (1994). According to the Musto-hypothesis (1987), suggesting that drug epidemics eventually die out when a new generation becomes aware of the dangers of drug abuse, we implemented an incidence scenario which includes the feedback effect of prevalence on incidence.

Under this scenario we are able to derive insights on the effectiveness of demand-sided controls, like prevention and treatment programs. It turned out that prevention will be the most appropriate means to reduce prevalence and consumption, but may exhaust more financial resources than necessary if applied over the entire epidemic. Additionally section 2.2 will show that treatment, when applied over the entire epidemic regardless of its course, might even intensify the problem. One would „treat away“ heavy users even at stages of the epidemic where they „would be needed“ as deterrent paradigm to non-users.

2.1 A model of the feedback effect of prevalence, or level, of use on initiation

We assume all initiates to start their habit as light users. Therefore at the beginning of the epidemic nearly all users are light ones. As time evolves some escalate to heavy use and the number of heavy users increases. Also the fraction of heavy users varies (until today it mainly increased) over time.

Initiation rates are significantly influenced by the current prevalence, or level, of use. In particular, most people who start using drugs do so through contact with a friend or sibling who is already using. Illicit drug use is „transmitted“ to non-users (very much like an infectious disease, hence the term „cocaine epidemic“) by drug consumers in the early stages of use. These users, who have not yet experienced the consequences of drug abuse, impress and „infect“ non-users with attracting descriptions full of enthusiasm. 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 hypothesizes that drug epidemics eventually die 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[1] to suppress rates of initiation into drug use.

Figure 2.1 depicts the deterrent effect of an increasing number of heavy users on initiation, while in absence of this effect initiation would grow with the number of light users. Peter Rydell suggested following endogenous function for the initiation into light use which describes all characteristic features of drug initiation discussed above.[2] The number of initiates per active light user per time unit

[1] According to Everingham and Rydell (p. 14, 1994): „Clinicians and researchers divide drug consumption into three levels: use (experimental, occasional, social consumption), abuse (regular, sporadically heavy, intensified consumption), and dependence (compulsive or addictive consumption)“. The last two categories correspond roughly to the group of heavy users.

[2] Note that we do not distinguish between incidence and relapse.

is determined by the reputation of cocaine *R(t)=exp [-q H(t)/L(t)]*, which is governed by the ratio of light and heavy users at time *t*. This function has the property to be invariant to proportional changes in *L* and *H*, where *L(t)* and *H(t)* denote the number of light and heavy users at time *t*, respectively. The initiation function (2.1) as well as the model (2.2) are designed for high prevalence, i.e. *L(t), H(t)>0*. Additionally we assume a group of non-users, „innovators" τ, to flow into the state of light use. τ does not depend on *L*, i.e. τ reflects the number of non-users who initiate on their own - for the sake of curiosity, shift from other drugs, or some other reason, but not through the urging of a friend who is cocaine user. Since all variables depend on the current stage of the epidemic we drop the explicit denotation of the dependence on time in what follows:

$$I(L,H) = \tau + sL\exp\left[-q\frac{H}{L}\right] \tag{2.1}$$

L>0 ... number of light users
H>0 ... number of heavy users
s=0.61 ... average rate at which light users attract non-users
q=7 ... constant measuring the deterrent effect of heavy drug abuse
τ *=50,000 ... number of innovators*

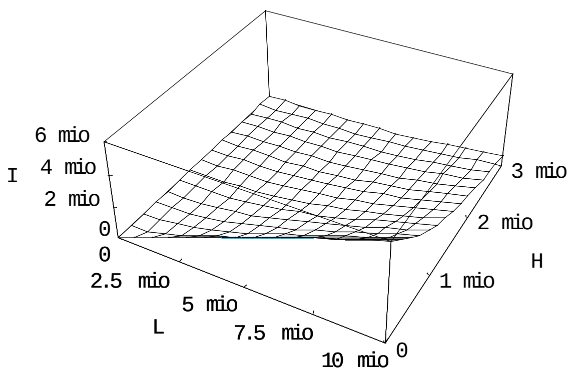

Figure 2.1

Three-dimensional plot of the initiation-function by P. Rydell
Parameter-set: *s=0.61, q=7.0,* τ *=50,000*

The values of the parameters *s* and *q* were found by minimizing the sum of the squared difference between modeled and observed annual cocaine initiation rates (see Behrens et al., 1997). The size of parameter *s=0.61* can be interpreted such that per year approximately two occasional users would „persuade" a nonuser to try cocaine, if there were no heavy users giving cocaine a bad reputation. The constant q which measures this deterrent effect corresponds roughly to the proportion at which the average heavy user consumes cocaine compared to the average light user.

The rest of our model is essentially a continuos-time analog of Everingham and Rydell's model. Due to the heterogeneity in the cocaine consumption behavior the US population can be roughly divided into three groups: non-users, light users and heavy users. Because the number of non-users is very large compared to the number of users it behaves like a constant which has not to be modeled explicitly (see Everingham, Rydell and Caulkins, 1995). The flow rates from one state to another are assumed to be proportional to the source states:

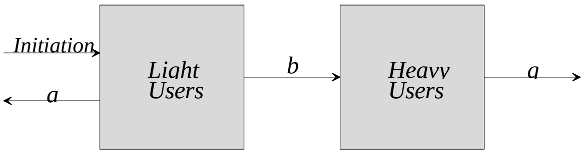

Figure 2.2

Flow diagram for model (2.2)

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 (1994, p.43). For example, they estimate that 15% of light users quit use each year and 2.4% escalate to heavy use, so we set $a = -\ln[1 - 0.15] @ 0.163$ and $b = -\ln[1 - 0.024] @ 0.024$, to three decimal places.

The one difference between this model and that of Everingham and Rydell is that Everingham and Rydell divided the outflow from heavy use into a flow out of use altogether (g) and a flow back into light use (labeled f). We drop 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 to heavy use than to enter the state for the first time. Hence we prefer to have only a flow from heavy use to non-use and view that rate as net of relapse.
Practically, Everingham and Rydell found that the data did not identify f and g individually. Any combinations of these parameters such that $f + g = 0.06$ fit the

historical data reasonably well. Since reducing the number of flows simplifies the analysis, we choose $f = 0$ and $g = -\ln[1 - 0.06] = 0.062$.

Now we can define a time-continuous dynamical Everingham & Rydell-type model for the evolution of cocaine consumption in terms of prevalence (number of light and heavy users) including the feedback effect of prevalence on initiation which is described by function (2.1):

$$\begin{aligned} \dot{L} &= I(L,H) - (a+b)L\,, \qquad & L(0)=L_0 \\ \dot{H} &= bL - gH\,, \qquad & H(0)=H_0 \end{aligned} \tag{2.2}$$

L>0	*number of light users*
H>0	*number of heavy users*
I(L,H)	*initiation into light use*
a@0.163	*average rate at which light users quit*
b@0.024	*average rate at which light use escalates to heavy use*
g@0.062	*average rate at which heavy users quit*

2.2 Do we „need“ heavy users at early stages of the epidemic?

Our analysis of a continuous analog to the Everingham and Rydell model of cocaine demand, augmented with an endogenous initiation function, generates a number of interesting observations which are explained in detail in the following subsections:

Proposition 2.1

For the base case parameter values ($a \cong 0.163, b \cong 0.024, g \cong 0.062, s = 0.61, q = 7.0,$ $p = 50{,}000$ *) the model follows a curve similar to that observed in historical data and approaches a steady state which is relatively small in size but different from zero. The number of steady state users is proportional to the number of innovators,* t*, who initiate of their own accord, not at the urging of current users, although the way the steady state is reached does not depend on* t.

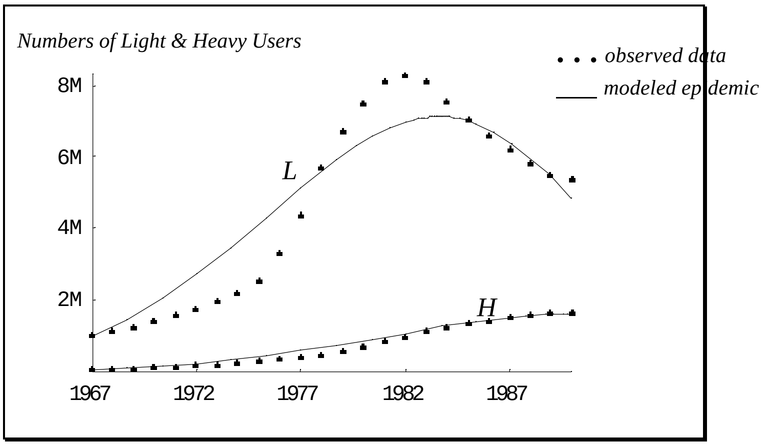

Figure 2.3

Time paths of the continuously modeled cocaine epidemic and the smoothed historical data (Everingham&Rydell)

Proposition 2.2

Damped oscillations always occur for flow rates within the limits given by Everingham and Rydell ($a \in [0.157, 0.168]$, $b \in [0.02, 0.03]$, $g \in [0.041, 0.094]$ *) and for a broad range of initiation function parameters (* $s \in [0.17, 1.0]$, $q \in [5.1, 11.5]$ *). The amplitude and frequency of the oscillations depend on the magnitude of the flow rates, especially on the rate at which light users escalate to heavy use, denoted by b.*

Proposition 2.3

In equilibrium about one-quarter of users are heavy users, and that proportion of heavy users deters most of the potential recruited initiation. [3]

Proposition 2.4

Reducing the rate at which light users attract non-users, s, and/or increasing the rate at which light users quit, a, which both can be affected by prevention programs, can reduce the size of the modeled cocaine epidemic.

[3] Just to avoid any misunderstanding: We do not claim for additional heavy users - and the problems and the costs caused by them. But their deterrent effect on incidence might be able to replace preventive programs at specific stages of the epidemic.

Proposition 2.5

Reducing the rate at which light users escalate to heavy use, b, and/or increasing the rate at which heavy users quit, g, increases the intensity of the modeled epidemic for initial data observed when the problem was arising.[4]

Proposition 2.6

For values of the rate at which heavy users quit, g, just slightly beyond the limits of the Everingham-Rydell interval (specifically at g_c *@0.094047) stable periodic solutions occur in a supercritical Hopf bifurcation. Deviations of this size are plausible and could be generated by treating large proportions of heavy users throughout the epidemic. Since such cycles are generally undesirable, this implies that treatment, though valuable at some times, should be targeted strategically to certain stages of a drug epidemic.*

It is clear from the observations above that the effectiveness of drug control interventions, notably treatment, vary over the course of an epidemic. Different strategies are most effective at different specific stages of an epidemic, and one expects the optimal mix of interventions to depend on the course and status of the epidemic (see chapters 3). Due to investigations based upon optimal control measures we may state that prevention programs are most effective in early stages of the epidemic, when most users are light users, while treatment programs are most effective when a greater fraction of users are heavy users, as is typical in the latter stages of an epidemic.

2.2.1 Effects of changes in flow rates on the equilibrium

In order to study the global qualitative properties of the solutions to system (2.2), we construct the phase diagram (Figure 2.4 and Figure 2.5) for the base

[4] Until the peak in light use appears in the current US cocaine epidemic, following happens: On the one hand the less light users escalate to heavy use - which seems to be desirable at the first sight - the more light users remain in their source state where they increasingly attract non-users. On the other hand if heavy use which is responsible for deterrence on incidence decreased, either by a cut back in the rate at which light users escalate to heavy use or by an increment in the rate at which heavy users quit, initiation and consequently the number of light users would grow and furthermore modeled consumption would almost „explode“ for a couple of decades.

case set of parameters $a \cong 0.163, b \cong 0.024, g \cong 0.062, s = 0.61, q = 7.0, t = 50{,}000$. The trajectories spiral counter-clockwise into a focus.

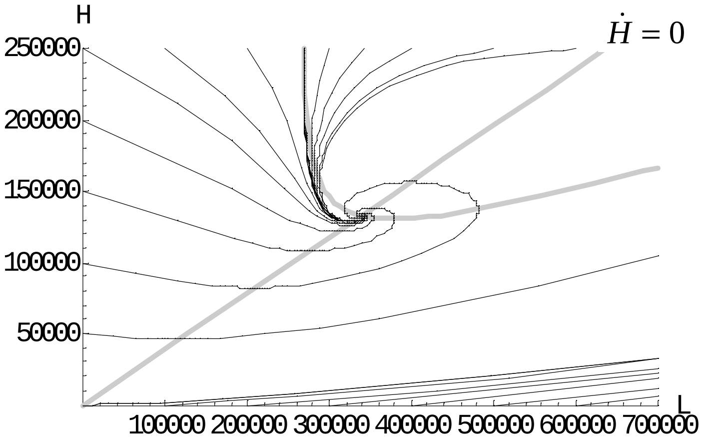

Figure 2.4

Phase portrait of model (2.2) for the base case parameter-set:
$a \cong 0.163, b \cong 0.024, g \cong 0.062, s = 0.61, q = 7.0, t = 50{,}000$. (Close-up of Figure 2.5.)

The locus of the points where the number of heavy users remains constant[5], is given by the constant ratio $H/L = 0.39$ and is depicted as the straight line $H = 0.39L$ in Figure 2.4 and Figure 2.5. For $H/L > 0.39$ the number of heavy users decreases and for $H/L < 0.39$ it increases. The curve which defines the locus where the number of light users remains unchanged is characterized by $H = -0.143L\ln\left[0.31 - \frac{81{,}967.2}{L}\right]$ and asymptotically approaches $L = 267{,}380$. Below and left to this $\dot{L} = 0$ isocline light use tends to increase and above this curve light use decreases. All trajectories starting in the positive quadrant spiral towards the unique equilibrium

$$\hat{E} = \begin{pmatrix} \hat{L} \\ \hat{H} \end{pmatrix} = \begin{pmatrix} 338{,}416 \\ 132{,}864 \end{pmatrix}.$$

This characterization of $\hat{E}$ as a stable focus can be confirmed locally by linearization of the differential equations (2.2) where we obtain a pair of conjugate complex eigenvalues with negative real parts, $l_{1,2} \cong -0.051 \pm 0.081i$. The amplitude

[5] For a general derivation of the formulas see Appendix A.2.1.

is determined by the size of the real parts of the characteristic roots and the frequency by the magnitude of the imaginary parts.

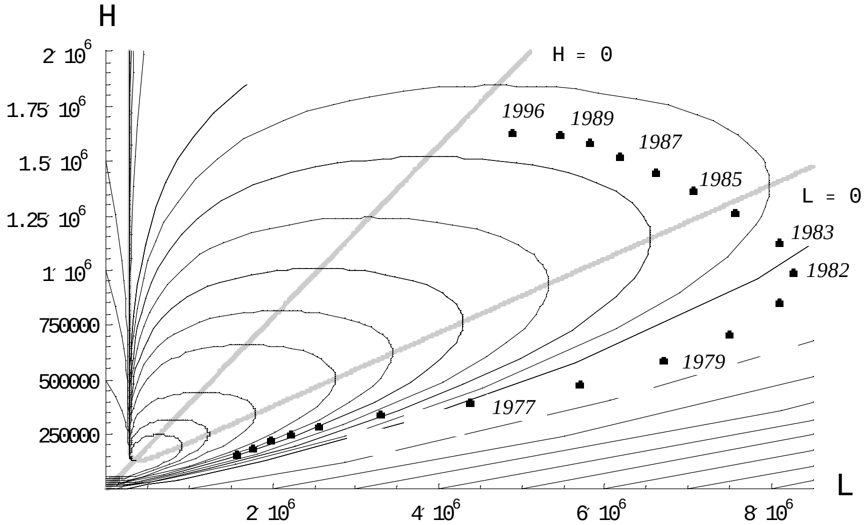

Figure 2.5

Phase portrait of model (2.2) for the base case parameter-set:
$a \cong 0.163, b \cong 0.024, g \cong 0.062, s = 0.61, q = 7.0, \tau = 50{,}000$, including smoothed historical trajectory of the current US cocaine epidemic (Everingham&Rydell)

Figure 2.5 also shows the historical evolution of the cocaine epidemic which to date has followed a spiral not so different from that which the model predicts. The modeled and historical data also move around their respective paths at nearly the same rate. The fit is not perfect; the historical data reflect a higher, sharper peak in light use. Nevertheless, the similarity is striking given that the actual epidemic has been 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 difference between historical and modeled data.

We obtain an equilibrium proportion of heavy users of $b/(b+g) = 0.28$, which is surprisingly large, taking into consideration that especially heavy users oppose very high costs to society. But that density of heavy users is sufficient to deter over 93% of potential recruited initiation. Just to avoid any misunderstanding, we do not claim that having so many heavy users – and the associated problems – is good. But their deterrent effect on initiation might help enhance or perhaps even replace preventive programs at certain stages of the epidemic.

2.2.2 Sensitivity analysis

Since the set of parameters estimated by Everingham and Rydell is not uniquely determined, we are interested whether the stability properties of system (2.2) remain invariant for variations in the size of the parameter values. According to Everingham and Rydell many combinations of values for a, b, g provided equally good fits to the historic data, though the range of these values is limited (see Table 2.1). Structurally we observe four behaviors: damped oscillation, exponential decay toward the equilibrium, stable limit cycles, and exponential growth. The last is not interesting because it does not match historical data and cannot continue indefinitely in a finite population.

Table 2.1: Effect of Parameter Values on System Behavior

Parameter	t	a	s	q	b	g
Base case value	50,000	0.163	0.61	7.0	0.024	0.062
E&R Range	NA	0.157 - 0.168	NA	NA	0.02 - 0.03	0.0 - 0.094
Range yielding exponential decay	None	0.348 - 1.0	0.0 - 0.166	>11.44	0.03791 - 1.0	0.0 - 0.04265
Range yielding damped oscillations	All	0.060255 - 0.348	0.166 - 1.036032	5.08 - 11.44	0.017965 - 0.03791	0.04265 - 0.094047
Range yielding stable limit cycles	None	0.021 - 0.060255	1.036032 - 1.6309	3.679 - 5.08	0.0134 - 0.017965	0.094047 - 0.138
Range yielding exponential growth	None	0.0 - 0.021	>1.6309	3.014 - 3.679	0.12 - 0.0134	0.138 - 0.1438
Hopf critical value	None	0.060255	1.036032	5.08	0.017965	0.094047

Stable limit cycles may emerge in a supercritical Hopf bifurcation if any of the parameters - except for the number of innovators, t - is deviated sufficiently from its base case value.[6] The Hopf theorem (see Appendix A.2.2) states that

[6] Note that we are able to exclude the case of subcritical Hopf bifurcations for each of the parameters. (see Appendix A.2.3).

limit cycles exist if (i) two purely imaginary eigenvalues exist for a critical value of the parameter, such that (ii) the imaginary axis is crossed with nonzero velocity. For *a* and *s* the Hopf bifurcation values could be calculated explicitly. For parameters *q*, *b* and *g* the conditions (Equations (A2.5) and (A2.6)) are given implicitly and must be calculated numerically.
Table 2.1 shows the ranges of the parameters for which Everingham and Rydell (1994, p.42) found good fits to the historical data and the ranges yielding each of the system behaviors. In general, the plausible parameter ranges are comfortably within the ranges which yield damped oscillatory behavior. The exception is parameter *g*. Section 2.2.3 discusses what happens if *g* begins to exceed its critical value.

Table 2.2: Effect of Parameter Values on Consumption

Parameter	t	*a*	*s*	*q*	*b*	*g*
Elasticity of $Q_{50}(1967)$	0.0057	- 0.0511	0.1012	- 0.0636	- 0.0527	0.0105
Elasticity of $Q_{50}(1975)$	0.0027	- 0.0255	0.0355	- 0.0324	- 0.0149	0.0023
Elasticity of $Q_{50}(1978)$	0.0013	- 0.0378	0.0664	- 0.0479	- 0.0308	0.0058
Elasticity of $Q_{50}(1981)$	0.0009	- 0.0264	0.0389	- 0.0336	- 0.0159	0.0023
Elasticity of $Q_{50}(1982)$	0.0008	- 0.0201	0.0257	- 0.0252	- 0.09	0.0003
Elasticity of $Q_{50}(1983)$	0.0007	- 0.0141	0.0146	- 0.0168	- 0.0034	- 0.0015
Elasticity of $Q_{50}(1985)$	0.0006	- 0.007	0.004	- 0.0065	0.0013	- 0.0036
Elasticity of $Q_{50}(1989)$	0.0006	- 0.0036	0.0007	- 0.007	0.0019	- 0.0047
Elasticity of $Q_{50}(1996)$	0.0006	- 0.0031	0.0004	- 0.0011	0.0019	- 0.0048
Elasticity[7] of equilibrium consumption $Q_{\hat{E}}$	0.01	- 0.0109	0.0027	- 0.0071	- 0.0042	0.00003

Concerning the sensitivity of consumption to parameter values, it is useful to consider both consumption in steady state and discounted total consumption

[7] $h_{Q_{\hat{E}}} = \dfrac{Q_{\hat{E}}(1.01*n) - Q_{\hat{E}}(n)}{Q_{\hat{E}}(n)}$, n denotes any of the parameters; $Q_{\hat{E}} = 16.42\hat{L} + 118.93\hat{H}$

over a relevant finite planning horizon such as 50 years, denoted by $Q_{50}(year)$[8], starting with various initial data from the current US cocaine epidemic. (Following Everingham and Rydell, we assume light and heavy users consume at a fixed annual rate of 16.42 grams and 118.93 grams per year, respectively, and discount at 4% per annum.) Table 2.2 reports the elasticity of those quantities with respect to each of the six parameters, where consumption appears less sensitive to parameter perturbations at the latter stages of the epidemic. The implications are as follows.

Ad* t*, s, and q

Immediately from Equation (A2.4) we can derive that the number of innovators t does not influence the stability properties of the equilibrium state. But its size increases and decreases, respectively, proportional to the number of innovators (see Equation (A2.3) in the appendix).

The effects of changes in, *s*, the rate at which light users attract non-users, can also be derived from the Equations (A2.3) and (A2.4). The equilibrium number of users declines with declining *s*. All values between *s@0.166* and *s@0.036* lead to damped oscillations, but the smaller *s* is, the sooner the epidemic peaks and the less intense the epidemic will be. This shows the importance of teaching non-users to resist peer pressure to use drugs. Especially at early stages of the epidemic cutbacks in *s*, as well as increments in *a*, which are both goals of prevention programs, can substantially reduce the magnitude of the cocaine epidemic.

Increasing *q*, which measures the deterrent effect of heavy users on initiation, has the expected effect of diminishing both the steady state level and the intensity of the epidemic. Variations of the parameter within the interval $5.08<q<11.4$ ensure damped oscillations. Since the values of *q* and *s* are linked (large values of *q* correspond to large values of *s*), large one-sided deviations from the base case parameter values are not considered.

Ad a

From Equation (A2.3) we can derive that a reduction in the rate at which light users quit, *a*, raises the steady state values, and according to Equation (A2.4) the

---

[8] $Q_{50}(year) = \int_0^{50} e^{-0.04t}\left(16.42L(t)+118.93H(t)\right)dt$ subject to model *(2)*, provided with initial data from the cocaine epidemic observed for the corresponding year.

amplitude and duration of the epidemic grow too. On the other hand it is obvious that, if more light users quit, the epidemic is less intense, shorter in time, and the stable state values lower in size. But only an outflow from the state of light use of more than *34.8%* leads to an exponential decay of the epidemic. Particularly in the beginning of the epidemic consumption is quite sensitive to changes in *a*. Even though the Everingham and Rydell interval is narrow (*0.157-0.168*), deviations within that range lead to deviations in total consumption over 50 years (starting with data from 1967) of up to approximately 21.5% from their base values. In contrast, steady state values change no more than 3.9%.

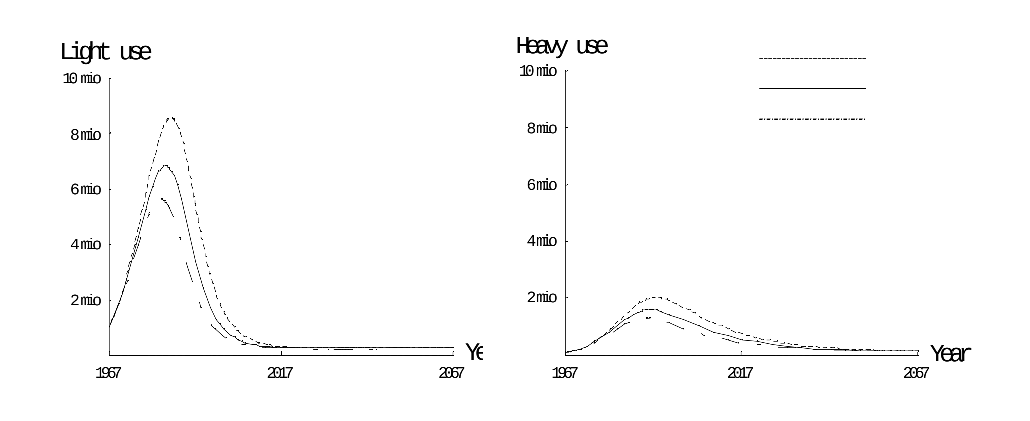

Figure 2.6

Time paths of light and heavy users for different values of the rate at which light users quit, *a*, for initial values from 1967

Ad b and g

For initial data from the beginning of the epidemic the model's behavior is very sensitive to changes in *b* and *g*, the two flow rates affecting heavy use. With respect to parameter *b*, reducing escalation to heavy use would at first seem to be good because heavy users consume at much higher rates than light users and persist longer in that use. On the other hand the fewer light users who escalate to heavy use, the more light users there are to attract non-users into use and the fewer heavy users there are to deter initiation.

If light users progress to heavy use rapidly (more than 3.7% of them per year), then the system approaches the steady state directly, without overshoot or oscillation. If very few light users progress to heavy use each year (less than

1.8%), then the system would not approach a single positive equilibrium. However, such values are outside the range Everingham and Rydell found to fit the historical data.

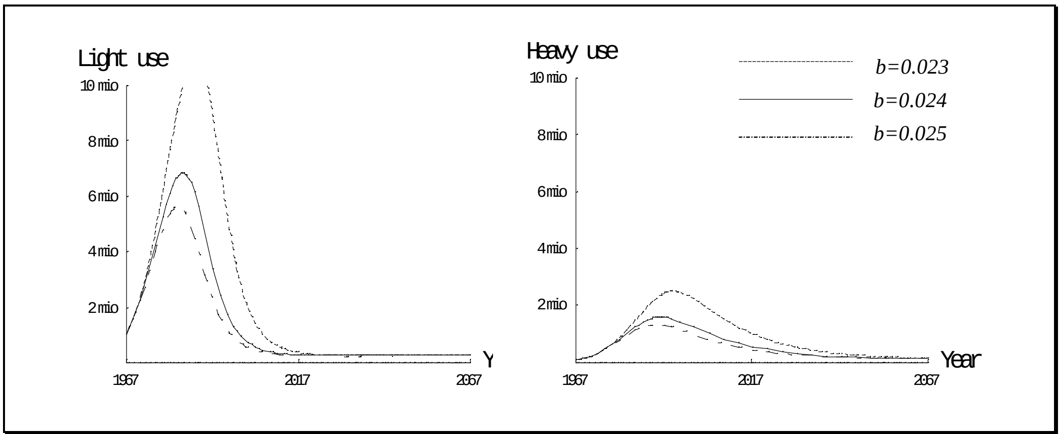

Figure 2.7

Time paths of light and heavy users for different values of the rate at which light users escalate to heavy use, *b*, for initial data from 1967

Even though variations in *b* do not threaten the basic system behavior, they do dramatically affect consumption for initial data from the beginning of the current cocaine epidemic. Figure 2.7 shows that small changes in the value of parameter *b* lead to large changes in the height of the peaks. If we increase *b* by *0.001* (dashed line in Figure 2.7), $Q_{50}(1967)$ is reduced by 14%. The corresponding rise in the steady state number of heavy users is 0.4%. This increases deterrence, reducing the equilibrium number of light users by 2.5%. On the other hand, if we decrease *b* by *0.001* (which is depicted by the dotted line in Figure 2.7) the equilibrium number of light users will increase by 5.3% while the equilibrium number of heavy users decreases by 0.3%. More dramatically, it leads to a 45% increase in total consumption over a 50 year time horizon because it tremendously increases the intensity of the arising epidemic.

Increasing the rate at which heavy users quit at early stages of the epidemic has the opposite effect that one would at first expect. Increasing *g* throughout the entire epidemic, which corresponds to treating a larger proportion of heavy users at all times, actually increases both the steady state number of heavy users and, even more dramatically, consumption over the first 50 years, for initial data observed in 1967. For example, if we increase *g* from *0.062* to *0.09*, there will be 56% more light users and about 7% more heavy users in steady state, and consumption over the first 50 years will be 121% higher. On the other hand, reducing the rate at which heavy users quit reduces consumption in the long run by increasing the number of heavy users who act as deterrent to initiation. E.g. decreasing *g* from *0.062* to *0.043* (dotted line in Figure 2.8) reduces the number of light users in steady state by 15.6%, increases the number of heavy users by 21.4%, and cuts consumption over the first 50 years by 20.5%.

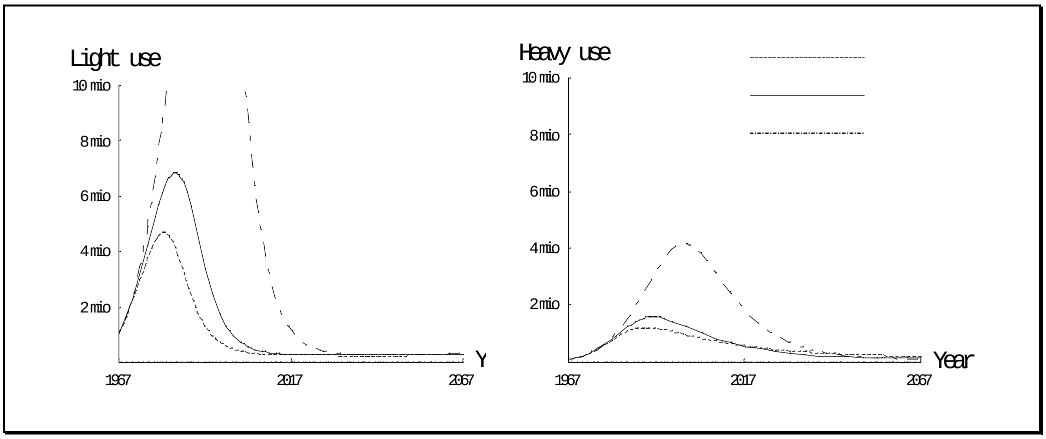

Figure 2.8

Time paths of light and heavy users for different values of the rate at which heavy users quit, *g*, for initial data from 1967

This by no means implies that treatment is never useful or that it has no role to play in drug control. Treatment at latter stages of the epidemic can help reduce drug consumption. The seemingly perverse result applies when one is obligated to pursue a given level of treatment throughout the entire course of the epidemic. Treating and removing heavy users when there are relatively few heavy users is counter-productive because each user contributes significantly to deterring initiation. If there are many heavy users relative to the number of light users, however, treating and removing some does not greatly jeopardize the negative reputation of the drug or the suppression of initiation.

Figure 2.9 makes this clear. It distinguishes combinations of numbers of light and heavy users for which additional treatment beyond the level historically pursued in the US (that which is reflected in the flow parameters) is counter-productive and combinations for which it is helpful. Within the region where additional treatment is helpful, the figure shows isoclines indicating combinations for which additionally treating a heavy user reduces consumption by a particular amount Δ. (For the calculation of Δ see Appendix A.2.4)

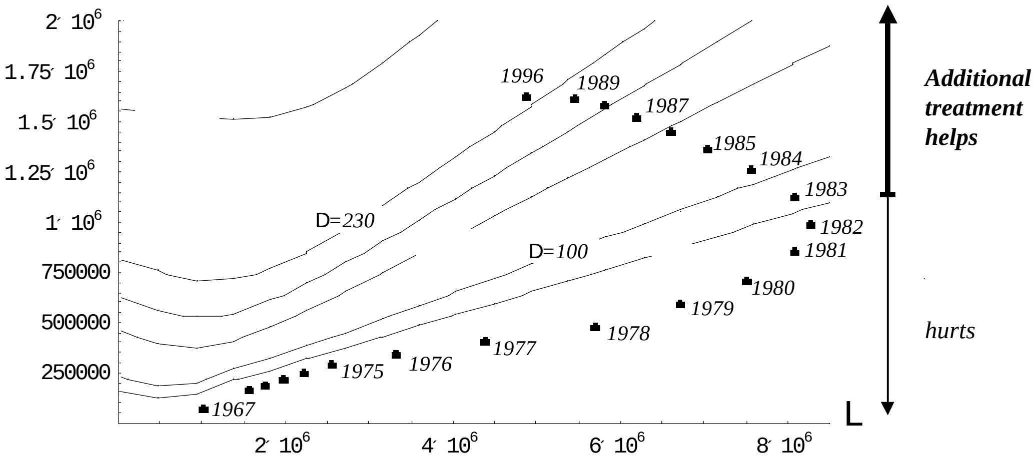

Figure 2.9

Contour plot for light and heavy users depicting isoclines for which additional treatment ($u = 31\% \rightarrow \tilde{u} = 39.2\%$ which implies $g = 6.2\% \rightarrow \tilde{g} = 7.2\%$) reduces consumption by Δ grams per additionally treated heavy user; Including smoothed historical trajectory of the current US cocaine epidemic (Everingham&Rydell)

The figure can be used as a guide to policy makers. If the epidemic is currently within the region where more treatment is counter-productive, then treatment levels should be below those typically observed in the US. If the epidemic is in the region where further treatment is effective, then the policy maker would have to estimate the cost of treating a user (perhaps around $2,000) and judge the benefit of reducing cocaine consumption by a certain amount. (The reduction in social cost per gram of consumption averted might be on the order of $100.) Together these figures determine how many grams of consumption a treatment must avert in order for additional treatment to be cost-effective. If the epidemic is above the corresponding isocline in Figure 2.9, then additional treatment is cost-effective.

2.2.3 Periodic prevalence

As mentioned above, for values of the rate at which heavy users quit in the interval $0.0427<g<0.094047$ we always find damped oscillations, but if g reaches 0.094947 a stable limit cycle emerges. Although the base value (0.062) is below the critical value (0.094047), for two reasons it is plausible that the actual value of that flow parameter might approach or exceed the critical value.

First, some people think the Everingham and Rydell values of g are too low (more precisely their values of $f+g$ are too low). They imply an average subsequent career length for a heavy user of $1/(f+g) = 15$ years or so, which some view as implausibly long. Second, the outflow rate g is affected by the proportion of heavy users who receive treatment. Everingham and Rydell assumed that historically about 31% of heavy users have received treatment each year and that 13% of those treated leave heavy use. Hence, the outflow rate can be viewed as being $g = 0.062 + 0.13(u - 0.31)$, where u is the proportion of heavy users receiving treatment each year. If that proportion increased to 55.7%, g would reach 0.094047.

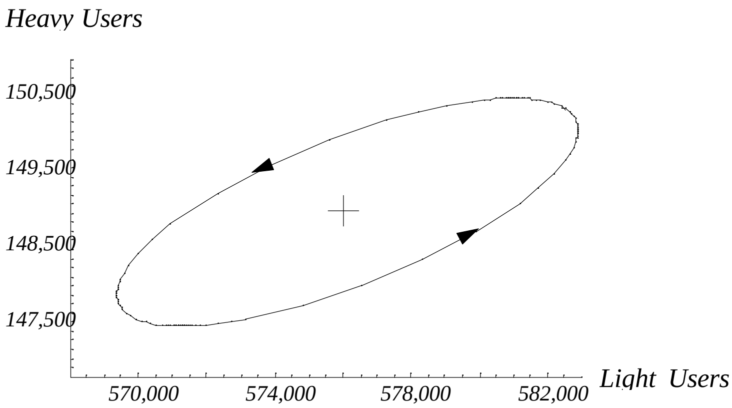

Figure 2.10

Cyclical prevalence depicted in the (L,H)-plane for *a@0.163, b@0.024, s=0.61, q=7.0,* t *=50,000 and g>g$_C$@0.094047*

When $g>g_C$, the unique equilibrium (Equation (A2.3)) changes its stability because the real parts of the characteristic roots (Equation (A2.4)) become positive and (A2.6) is larger than zero, *D@1.24*, implying that the eigenvalues cross the imaginary axis with positive velocity. We obtain a supercritical Hopf bifurcation (see Appendix A.2.3), where the orbits emerge for $g>g_C$. The fixed points $\hat{E}(g)$

have the local property of unstable foci and the periodic orbits are attracting. Hence, a deviation beyond the limits of the Everingham-Rydell interval could result in periodic prevalence. Note that close to the bifurcation point the cycles „grow“ with the size of $\sqrt{g-0.094047}$, until g reaches *0.138,* which corresponds to the hypothetical case of 89.5% of heavy users receiving treatment each year. (For even larger proportions of heavy users receiving treatment the cycles disappeared and the modeled system dynamics converged to infinity.)

Figure 2.11

Cyclical prevalence for parameter set $a \cong 0.163$, $b \cong 0.024$, $s=0.61$, $q=7.0$, $\tau=50{,}000$
and for different values of the rate at which heavy users quit,
$g_C \cong 0.094047<g<0.099$

As Figure 2.10, 2.11, and 2.12 depict, for a flow rate only infinitesimally larger than *0.094047* the numbers of light and heavy users would cycle counter-clockwise forever with a period of approximately 70 years (according to Equation (A2.9)). Since such cycling in a social system is usually undesirable, this is further evidence that treating a constant, large proportion of heavy users throughout an epidemic regardless of its course might not be the best policy.

Figure 2.12

Cyclical prevalence for parameter set *a@0.163, b@0.024, s=0.61, q=7.0,* t *=50,000*
and for different values of the rate at which heavy users quit,
g_C*@0.094047<g<0.094065*

Chapter 3

Drug Control Policies

The drug control budget nowadays exceeds \$15 billion - where the main part goes to different enforcement programs. (e.g. according to Rydell and Everingham, 1994, p.44: during 1992 \$0.9 billion were spent on source country control, \$1.7 billion on interdiction and \$9.5 billion on domestic enforcement). Treatment programs received „only" \$0.928 billion in 1992 and this sum already covers governmental as well as private money. But Rydell, Caulkins and Everingham (1996) showed that treatment emerges as the most cost-effective program compared to different enforcement strategies and Rydell (1997, p.21) showed that prevention is by far the most cost effective program. For almost the entire first half of the epidemic prevention would dominate both treatment and enforcement, and it would dominate enforcement until the latter stages of the epidemic. \$0.25 billion spent on school-based prevention would cut down initiation by more than 7% (Rydell, 1997, p.8).

Therefore the controls regarded in following optimization models are prevention and treatment, while enforcement is exogenously determined and treated outside the model. Generally speaking, we model a continuous-time optimization problem where a decision maker seeks to minimize the discounted stream of the sum of social costs caused by illicit drug use and the monetary costs due to control measures, so the utility functional, J, is given by

$$J = -\int_0^{\infty} e^{-r\,t} \left(\mathrm{k}\; Q(t) + u(t) + w(t) \right) dt \qquad (3.1)$$

where the constant k represents the social cost per gram of consumption, $Q(t) = 16.42\, L(t) + 118.93\, H(t)$, and $u(t)$ and $w(t)$ denote the expenditures on treatment and prevention programs at time *t*, respectively. The infinite time horizon stresses out that one is interested in regularities that are not caused by electoral cycles. $r > 0$ is the discount (time preference) rate. If *r* is small (close to zero), then the government is interested in what happens on the long run („farsighted"), whereas a „myopic" government is characterized by rather big values of *r*. The government's aim is defined by

$$\max_{u,w} J\,. \qquad (3.2)$$

Treatment as well as prevention are demand controls which reduce cocaine consumption directly by lowering the number of users and by reducing the amount of consumption while patients are in treatment. 73% of addicts in outpatient programs do not consume any cocaine during treatment, as well as 99% of the residential patients. An average amount of $1,700 - $2,000 spent on a heavy user's treatment provides a 13% chance to exit addiction. In following models the percentage of heavy users who exit their source state at time *t* due to treatment will be denoted by b as depicted in Figure 3.1.

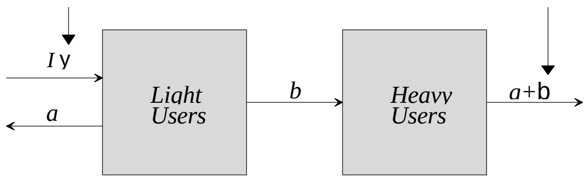

Figure 3.1

Flow diagram for the optimization models, where the measures of control are either prevention from light use or treatment of heavy use

Primary prevention is able to cut down initiation to a certain percentage y of its „uncontrolled" value. In mathematical terms, *I (1*-y*)* non-users will never enter light use though they were endangered before they received intense motivation to resist drugs, and information on how to do so. This includes e.g. school-based prevention, after school programs, clubs, broadcast media advertisements, etc. Prevention spending as well as expenditures on treatment have a diminishing effect on additional resources devoted to these programs. We model the efficiency of prevention spending $\psi(w)$ by a simple function with an exponential decay approaching an asymptotic value of *h*:

$$\psi(w) = h + k \exp[-mw] \tag{3.3}$$

h ... min. percentage to which initiation can be cut down due to prevention
k ... max percentage by which initiation can be reduced
m ... constant measuring efficiency of prevention

About 16% reduction in initiation is the best achievable result and this result is already obtained for approximately \$3 billion prevention spending per year. Spending more than that was a waste of resources and therefore the marginal efficiency of prevention is declining, i.e. $\psi' \leq 0$.

The valuation of the efficiency of treatment is a bit more pretentious. It is plausible that treatment has a diminishing effect, because the more heavy users are treated the more difficult it will get to involve further addicts into a treatment program and finally one would wind up with users who don't want to be treated at all. Each extra dollar spent in treatment has less impact than the one spent before. Therefore we choose a monotonically increasing function which is concave in expenditures on treatment per heavy user, i.e. $\beta'>0$, $\beta''<0$. The constant δ can be interpreted as a share for administrative cost

$$\beta(H,u) = c\left(\frac{u}{H+\delta}\right)^{d} \tag{3.4}$$

c ... constant measuring the efficiency of treatment
d ...constant measuring the marginal efficiency of treatment

Ideally the government would be free to choose whatever controls *u* and *w* minimized the objective function. In practice, however, the political process and the limitation of human institutions are such that controls cannot always be pursued. One simplistic image is that society allocates resources proportional to the magnitude of the problem, i.e.

$$u(t)+w(t)=g\,\Gamma\big(L(t),H(t)\big) \qquad (3.5)$$

When the drug problem is perceived to be small, the public might resist having many tax dollars being spent on drug control; they might prefer that scarce resources be allocated to more pressing issues. On the other hand, when the drug problem is perceived to be severe, the public might demand that significant interventions are made. Since the quantity or weight consumed has advantages as a general purpose scalar measure of the size of the problem (Rydell et al., 1996) we assume the drug control budget to be proportional to total consumption[9], i.e.

$$\Gamma\big(L(t),H(t)\big)=Q(t)=16.42\;L(t)+118.93\;H(t) \qquad (3.6)$$

In the present section we will investigate optimization models governed by three different budget rules: (a) constrained budget that is divided in a fixed proportion between treatment and prevention for all times, (b) constrained budget that can be allocated optimally between treatment and prevention, and (c) the completely unconstrained optimal control problem. These models are parameterized with data from the US cocaine epidemic because of its magnitude and because data on that epidemic are relatively good, but it is at least plausible that the qualitative conclusions generalize to similar drugs in similar contexts. (The base case parameter values were chosen in consideration of Caulkins, 1996; the report by the Office of National Drug Control Strategy, 1996; Rydell et al.,1996; and Rydell&Everingham, 1994)

---

[9] The parameter g represents the proportion for drug control spending. According to the *1994 Office of National Drug Control Strategy Report*, actual 1993 spending at the „federal" level on prevention for all drugs was \$1.5565 billion versus \$2.339 billion for treatment. This suggests that cocaine's share of the national prevention and treatment programs might be \$1.545 billion (assuming proportionalities between national and federal and treatment and prevention), so that g might be *5.31*.

Table 3.1: Base case parameter values

Parameter	Value	Description
a	0.163	rate at which light users quit their habit
b	0.024	rate at which light users escalate to heavy use
g	0.062	rate at which heavy users quit abuse or dependence
τ	50,000	number of innovators
s	0.61	rate at which light users attract non-users
q	7.0	constant which measures the deterrent effect of heavy use
c	0.0005	constant measuring the valuation of the efficiency of treatment
d	0.6	constants measuring the marginal efficiency of treatment
δ	0.00001	administration constant
h	0.84	min. percentage to which initiation can be cut down due to prevention
k	0.16	maximum percentage by which incidence can be cut back
m	2.37×10^{-9}	constant measuring efficiency of prevention
r	0.04	discount rate
γ	5.31	budget proportionality constant
κ	113	consumption proportionality constant

3.1 Constant fraction budget allocation

As an introduction to what nonlinear control programs can achieve we start our investigation with a simple, easy to implement, but sub-optimal, budget rule. This policy has the advantage that implementation cost - which is neglected in all models presented here - is comparatively low. Suppose one constraint the budget not only so that $u(t) + w(t) = Q(t)$, but also so that $u(t) = f\gamma Q(t)$ and $w(t) = (1-f)\gamma Q(t)$. I.e. f is the proportion of the budget going to treatment and $(1-f)$ is the proportion going to prevention, throughout the entire epidemic regardless of its course. Consumption is modeled by Equation (3.6) and the constraints are extensions of model (2.2) according to Figure 3.1. Initiation is modeled by function (2.1), prevention incorporated according to function (3.3) and

treatment according to function (3.4). From now on we omit the explicit denotation of time.
The government may wish to choose $0 \le f \le 1$, optimally once and for all, i.e. such that the cost of control and the societal cost, both roughly proportional to the quantity of cocaine consumed nation-wide, is minimized.

$$f^* := f^*\left(L_0, H_0\right) = \arg\max_f \; J_1$$

with

$$J_1 := J_1\left(f\left(L_0, H_0\right)\right) = -\left(\kappa + \gamma\right)\int_0^\infty e^{-rt}\, Q\, dt \tag{3.7}$$

subject to

$$\dot{L} = I\left(L, H\right)\psi\left(\left(1 - f\right)\gamma\, Q\right) - \left(a + b\right) L\,, \qquad L(0) = L_0$$

$$\dot{H} = b\, L - \left(g + \beta\left(H,\; f\gamma\, Q\right)\right) H\,, \qquad H(0) = H_0$$

This model might be appropriate if bureaucratic interests in treatment and prevention agencies were so strong that it is very difficult to vary the mix of treatment and prevention over time. Before dismissing such a constraint as artificial, note that treatment's share of the federal drug control budget in the US (including treatment research) was never less than 18.4% nor more than 22.3% between 1987 and 1997 (ONDCP, 1996).

Because the constant fraction model turned out to be by far not so efficient for the determination of optimal control policies for the entire epidemic than the model with constrained budget that can be allocated optimally between treatment and prevention, and particularly the completely unconstrained optimal control problem, we omit the description of the analysis of system (3.7) and list the most interesting observations done for the base case set of parameter values (Table 3.1) in the following subsections. The consequences for decision making, however, will be content of chapter 4.

3.1.1 Stability properties of the equilibrium

The optimal constant budget allocation, $f^*(L_0, H_0)$, depends on the size of the initial conditions, (L_0, H_0). For the base case parameter values (Table 3.1) the stability behavior of the equilibrium state, $\hat{E}_1 = (\hat{L}_1, \hat{H}_1)$, of nonlinear control model (3.7) can be classified with the help of stability regions. (See Figure 3.2.)

- The system dynamics approach a stable focus for initial values from region *Ia* and *Ib*. Data located in region *Ia* lead to full scale prevention programs, while values from region *Ib* result in a constant mix of prevention and treatment funding.

- The system dynamics approach a limit cycle for initial values from region *IIIa* and *IIIb*. Region *IIIb* corresponds to full scale treatment, whereas region *IIIa* includes data yielding a 96% - 100% treatment-share of the drug control budget.

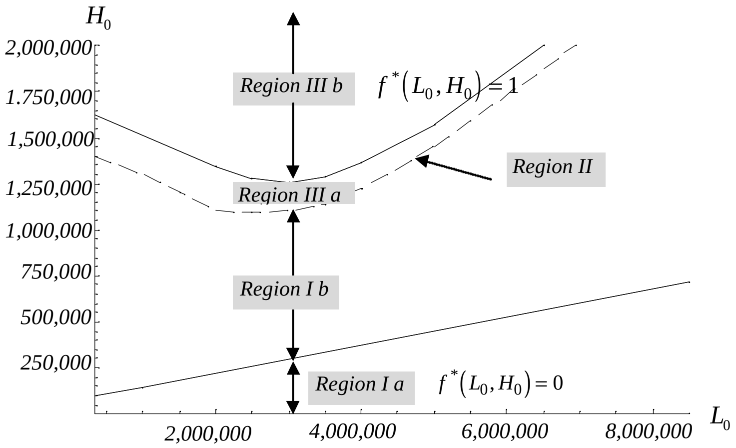

Figure 3.2

Stability regions of system (3.7) for initial combinations of light and heavy users

For initial values from region *II* ($f^*(L_0, H_0)_c \cong 0.95963352$) limit cycles emerge due to a supercritical Hopf bifurcation (see Appendix A.2.2). Treatment holding roughly 96% of the drug control budget at all times leads to cycling levels of use (with a period of approximately 71 years). This behavior corresponds to periodic prevalence for a proportion of heavy users receiving treatment of more than 55.7% in the framework of model (2.2).

The implementation of a constant budget allocation policy improves the size of the drug problem, in terms of steady state consumption, for initial data located below the $f^*(L_0, H_0) = 0.25$ isocline. As Figure 3.3 depicts e.g. steady state consumption is suppressed to 95% of its uncontrolled value for initial conditions from region *Ia* (see Figure 3.2). For a higher proportion than a 25% treatment-share of the budget the problem would even be intensified. Similar to model (2.2) treatment measures were applied to the epidemic even at stages where heavy users were „needed“ to deter and suppress initiation. Note that model (2.2) is a special case of the constant fraction allocation problem (3.7) with $g=0$. Therefore the similarities in system behavior are not really surprising.

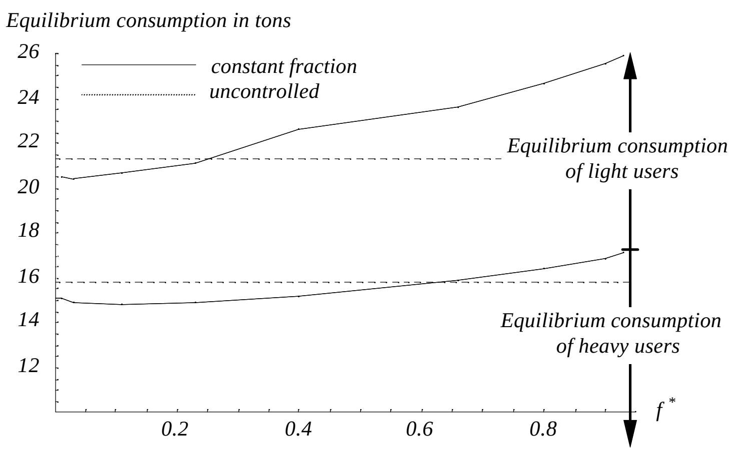

Figure 3.3

Budget allocation proportion $f^*(L_0, H_0)$ plotted vs. equilibrium consumption for initial data from region *Ia* and *Ib* (see Figure 3.2); including uncontrolled steady state consumption

If all resources are invested into prevention programs, one observes exactly the same equilibrium fraction for model (3.7) as for model (2.2), but equilibrium consumption is cut back by 5% compared to its uncontrolled value. Furthermore, contrary to equilibrium consumption, the equilibrium fraction of heavy users decreases with a growing share of the budget for treatment programs. (See Figure 3.4.) Due to the feedback effect of prevalence on initiation the number of users, and, consequently, equilibrium consumption, increases if larger proportions of heavy users are removed by treatment programs at „too early" stages of the epidemic.

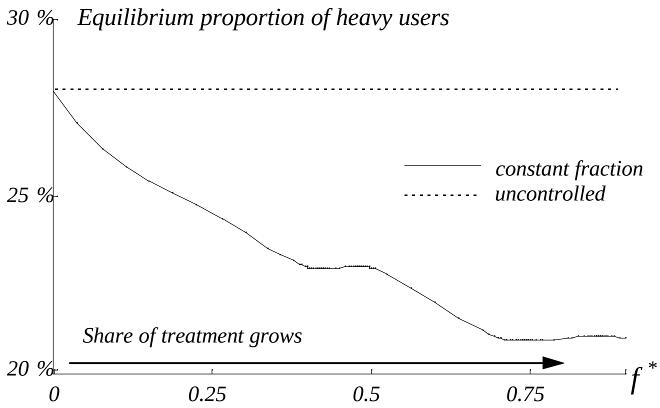

Figure 3.4

Budget allocation proportion $f^*(L_0, H_0)$ plotted vs. the equilibrium fraction of heavy users for initial data from region *Ia* and *Ib* (see Figure 3.2)

3.1.2 Observations on the constant fraction budget for initial data from the current US cocaine epidemic

As motivation for the analysis of optimal control model (3.8), where the allocation of the budget may be altered over time, we are particularly interested in the change of the optimal constant fraction, $f^*(L_0, H_0)$, for different stages of the current US cocaine epidemic (see Figure 3.5) and the effects of these alterations on the value of the objective functional, and, consequently, on total costs. These costs are calculated as the sum of the societal cost from 1971 until the year of control's implementation and the corresponding values of the maximized objective functional, $J_1^* := J_1\left(f^*(L_0, H_0)\right)$.

The Everingham&Rydell-estimates of the current cocaine epidemic were used as initial conditions for calculations done by means of model (3.7) yielding time-discrete values of the optimal fraction, $f^*(L_0, H_0)$, for each year. Implications are as follows.

Observation 3.1a

Under the assumption that the budget is ruled by a constant fraction policy the decision-maker's preference might have

- *gone entirely to preventive measures for data from the beginning of the current US cocaine epidemic.*
- *shifted from prevention toward treatment for initial data from the 1980s.*
- *been entirely devoted to treatment programs for data observed since 1991 (though this strategy results in periodic prevalence).*

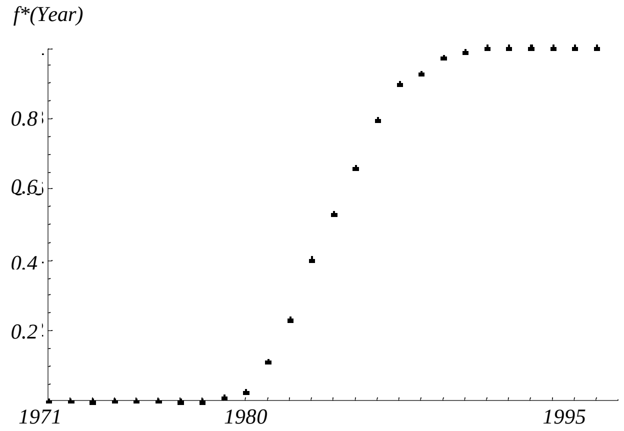

Figure 3.5

Optimal fractions $f^*(L_0, H_0)$ calculated for initial data, (L_0, H_0), from the years 1971-1996.

In 1976 light use started to grow increasingly, but the analysis of model (3.7) suggests that until 1978 the best choice of the budget allocation would have been to devote the entire financial resources to prevention programs[10], though maintenance of this strategy would have implied increasing cost (see Figure 3.6). Up until 1982, where the number of light users peaked, prevention would have covered always more than ¾ of the budget, whereas it would have already been left with a 1% share in 1990 (see Figure 3.5). Ever after 1991 full scale treatment programs would have been able to minimize the objective functional.

[10] Note that there is actually a six-year lag between spending money on prevention and seeing its effect on initiation. Adding such a delay to the model might alter the timing of the „shift" from prevention spending toward treatment spending.

Observation 3.1b

The earlier one implements the constant fraction budget control, for data observed during the current cocaine epidemic, the lower the total costs.

If we compare total costs, i.e. the sum of the societal cost from 1971 until the year of control's implementation and the corresponding value of the maximized objective functional ever after control's implementation, a 1971 full scale prevention program would have drastically reduced costs, i.e. to 31% of their present value. The „1983-plan“ (proposing 40% of the budget for treatment programs) would have already implied total costs of 75% of their present value (see Figure 3.6). The costs grow more or less with the time which elapses until control measures are implemented.

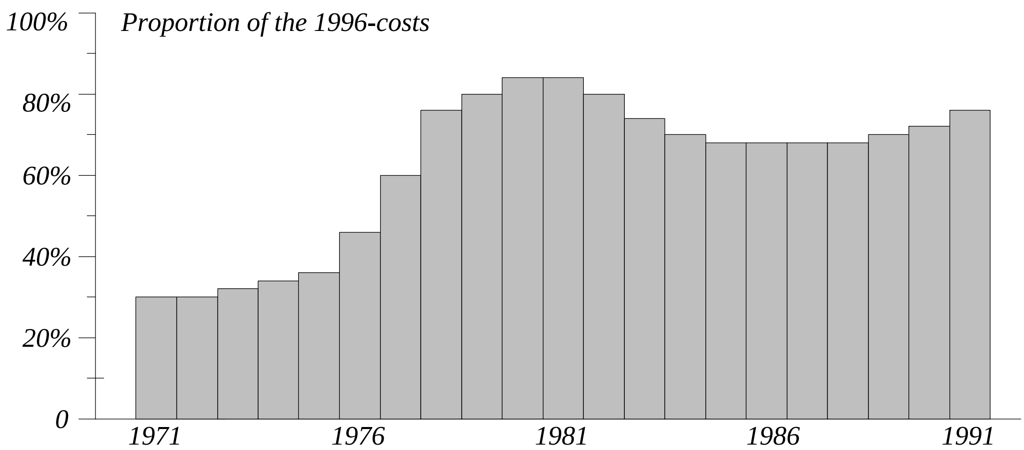

Figure 3.6
Proportion of present costs, measured in constant years dollars, for the years 1971-1991

Since according to model (3.7) all resources ought to be devoted to treatment nowadays, treatment expenditures would amount approximately $1.34 billion for the year 1997 (which is less than the present budget for preventive and treatment measures) - but increases equilibrium consumption above its uncontrolled value, and results in periodic prevalence according to the constant fraction budget rule.

Observation 3.1c

The cost of failure in the choice of the optimal budget allocation would have been most expensive for data from the beginning of the epidemic. Data observed during the 1980s, and at the present stage of the epidemic lead to reasonable errors. („The less you can win, the less you can loose.“)

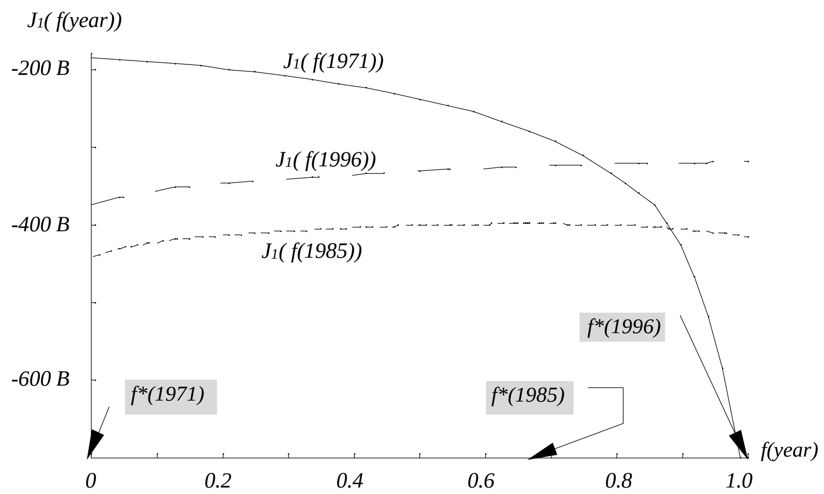

Figure 3.7
Objective functional as a function of the budget allocation proportion *f* for initial data observed in 1971, 1985, and 1996

If someone decided in 1971 that treatment received only a 1% share of the budget instead of nothing, costs would have grown about $993 million, because treatment was applied „too early“. Contrary, a 1% deviation from the present full scale treatment policy would cost „only“ $26 million. For initial data from the year 1985, where light use peaked already, the course of the objective functional as a function of *f*, $J_1\left(f(year)\right)$, peaks at $f^*(1985) = 0.66$. Increasing prevention’s proportion of the budget from 34% to 35% would have increased cost by $17 million.

We will use the results of the model presented here to compare efficiency and cost of the constant budget allocation policy with the effectiveness of a policy where one was free to alter the proportion of allocation over time in what follows.

3.2 Optimal mix of prevention and treatment

The government's objective is assumed to be the minimization of the cost of control and the cost to society which both are assumed to be proportional to the quantity of nation-wide consumed cocaine, where $Q=16.42L+118.93H$ denotes the quantity totally consumed. I.e. we consider the nonlinear control problem (3.8) with a single[11] scalar control u and two states, L and H, where the current value, using the discount rate $r>0$, of total consumption has to be minimized. Initiation is described by function (2.1), prevention is incorporated according to function (3.1) and treatment according to function (3.2).

$$u^* = \arg \max_{0 \le u \le \gamma Q} J_2$$

with

$$J_2 = -(\kappa + \gamma)\int_0^\infty e^{-rt}\, Q\, dt \tag{3.8}$$

subject to

$$\dot{L} = I(L,H)\,\psi\,(\gamma Q - u) - (a+b)\,L\,, \qquad L(0) = L_0$$
$$\dot{H} = b\,L - (g + \beta\,(H,u))\,H\,, \qquad H(0) = H_0$$

The analysis of the optimal control problem (3.8) proceeds in the usual manner (for the general theory see e.g. Leonard and Long, 1992; Feichtinger and Hartl, 1986), where $\mathcal{H}$ denotes the Hamiltonian and π_1 and π_2 denote the costate variables, all in current values

$$\mathcal{H} = -(\kappa + \gamma)(16.42L + 118.93H) + \pi_1\left(I\psi - (a+b)L\right) + \pi_2\left(bL - (g+\beta)H\right) \tag{3.9}$$

Applying Pontryagin's maximum principle to (3.9), we derive the necessary condition for an optimal interior control $u*$

$$\mathcal{H}_u = \pi_1 I\psi_u - \pi_2 H\beta_u = 0\,. \tag{3.10}$$

[11] Due to the definition of the budget rule, we can replace prevention by $w = \gamma Q - u$

A sufficient condition for the Hamiltonian $\mathcal{H}$ to be concave with respect to the control u is that the costate π_1 is negative (see Appendix A.3.1). Since $\pi_1<0$ has to hold on the stable manifold $u*$ is truly a maximum.

Condition (3.10) allows us to express the optimal interior control $u*$ implicitly in terms of states and costates, which leads to the so-called canonical equations (Equations (A3.3)). A result of Dockner (1985) allows us to characterize the local stability behavior of the equilibrium state of the canonical system (see Appendix A.3.2). The central question is to characterize the intertemporal equilibrium strategy, whether it is constant or cyclical, stable or unstable.

3.2.1 Equilibrium consumption and parameter sensitivity

For the base case parameter values, listed in Table 3.1, the equilibrium values for states, costates, and controls as well as the corresponding eigenvalues of the linearized system in the neighborhood of the steady state are given below:

$$\hat{E}_2 = \begin{pmatrix} \hat{L} \\ \hat{H} \\ \hat{\pi}_1 \\ \hat{\pi}_2 \\ \hat{u}* \\ \hat{w}* \end{pmatrix} = \begin{pmatrix} 347{,}037 \\ 124{,}646 \\ -36{,}842 \\ -12{,}576 \\ 7{,}344{,}350 \\ 101{,}630{,}000 \end{pmatrix} \Rightarrow \begin{cases} \lambda_{1,2} \cong \ \ 0.0828 \pm 0.0892\, i \\ \lambda_{3,4} \cong -0.0426 \pm 0.0892\, i \end{cases}$$

Due to the conjugate complex eigenvalues transient oscillations will always occur. Since two and only two eigenvalues have negative real parts, $\hat{E}_2$ has the properties of a saddle focus, i.e., for given initial states, choosing the corresponding initial conditions for the costates from the stable two-dimensional manifold ensures convergence to the equilibrium, while all other initial conditions of the costates lead to divergence. Therefore each pair of initial values (L_0, H_0) defines a unique trajectory on the stable manifold, which is the optimal solution of system (3.8).

Compared to the uncontrolled epidemic (model (2.2)) equilibrium prevalence of model (3.8) remains approximately equal in size, but the equilibrium fraction of heavy users, $\hat{H}/(\hat{L}+\hat{H})$, decreases by 7%. Therefore, equilibrium consumption,

$Q_{\hat{E}_2}$, decreases by 3.6%. (See Table 3.2.) This effect is achieved by a steady state prevention, $\hat{w}$, which amounts 2/5 of what Rydell (1997) claimed for in order to implement a nationwide school-based prevention program, and treatment, $\hat{u}$, reduces to $7.3 million at the steady state.

In the current cocaine epidemic, about 1982, a rapid growth in the number of heavy users occurred. The ratio $L : H$ fell from $10 : 1$ to $5 : 1$ in just five years and moved furthermore down to $4 : 1$ and this changed the nature of the cocaine epidemic drastically. The equilibrium ratio, $\hat{L} : \hat{H} = 2.8 : 1$ is an improvement compared to the last years, where it even fell down to $1.8 : 1$ in 1992 and 1993 (*The National Drug Control Strategy: 1996*).

Table 3.2: Equilibrium quantities of model (3.8)

$Q_{\hat{E}_2}$	$\frac{\hat{H}}{\hat{L}+\hat{H}}$	$\frac{\hat{u}^*}{\hat{w}^*+\hat{u}^*}$
20,522,496 grams	26%	6.7%

Table 3.3 summarizes the effects of 1% changes from the base case parameter values (Table 3.1) on the steady state, equilibrium consumption, the fraction of treatment of the budget, as well as the fraction of heavy users. t is roughly proportional to the steady state (comparable to Equation (A2.3) for model (2.2)) and equilibrium consumption, while the fraction of heavy users is not affected at all. Apart from this the numbers of light and heavy users and equilibrium consumption react only to changes in the rate at which light users quit, *a*. Prevention is quite sensitive to all flow rates affecting light use, *a* and *b*, as well as to the measure of deterrence, *q*. The effects of 1%-deviations from the base case parameter values on treatment and the share of treatment, however, dominate by far the effects on all other equilibrium quantities. Especially deviations in the flow rates affecting heavy use, *b* and *g*, and the measure of deterrence, *q*, lead to more than 12% changes in equilibrium treatment. The marginal efficiency of treatment, *d*, affects nothing else but size and proportion of equilibrium treatment. *h*, t and g just slightly affect the equilibrium quantities, except that g is proportional to equilibrium prevention.

Table 3.3: Effects of a 1% shortage of the base case parameter values on the equilibrium state $\hat{E}_2$, equilibrium consumption $Q_{\hat{E}_2}$, fraction of heavy users $\hat{H}/(\hat{L}+\hat{H})$, and treatment's share of the budget.[12]

-1%	$\hat{L}$	$\hat{H}$	$\frac{\hat{H}}{\hat{L}+\hat{H}}$	$Q_{\hat{E}_2}$	$\hat{u}^*$	$\hat{w}^*$	$\frac{\hat{u}^*}{\hat{w}^*+\hat{u}^*}$
a	+ 0.87%	+ 1.16%	+ 0.2%	+ 1.1%	- 4.42%	+ 1.47%	- 5.43%
b	+ 0.29%	+ 0.1%	- 0.14%	+ 0.15%	- 15.3%	+ 1.26%	- 15.41%
g	- 0.12%	+ 0.07%	+ 0.14%	+ 0.02%	+ 14.65%	- 1.04%	+ 15.61%
t	- 0.95%	- 0.96%	0%	- 0.95%	- 0.83%	- 0.97%	+ 0.12%
s	+ 0.03%	- 0.36%	- 0.29%	- 0.25%	+ 7.52%	- 0.82%	+ 7.78%
q	+ 0.2%	+ 0.9%	+ 0.51%	+ 0.7%	- 12.31%	+ 1.64%	- 12.9%
d	+ 0.05%	- 0.02%	- 0.05%	0%	- 2.18%	+ 0.16%	- 2.18%
h	+ 0.04%	+ 0.05%	+ 0.01%	+ 0.06%	- 0.13%	+ 0.06%	- 0.17%
g	+ 0.03%	+ 0.05%	+ 0.01%	+ 0.04%	- 0.33%	- 1%	- 0.64%

***Sensitivity with respect to social cost per gram,* k*, budget per gram,* g**

Sensitivity analysis with respect to the social cost per gram, k, is particularly important because these costs are not only difficult to measure, they are also inherently subjective because different people may wish to include or exclude different costs in a social planner's objective function. E.g., Rydell and Everingham (1994, p.38) report that societal cost estimates (based on Rice et al., 1990) of $19.68 billion for cocaine in 1992, which is associated with 291 metric tons of consumption or k*=67.6*. Rice et al. place very low estimates on the costs associated with crime (e.g. no pain or suffering costs), so another reasonable number would be roughly twice as large, k*=113*. A high end number would be about three times as large (saying that 60% of a $100 billion/year social cost could be attributed to cocaine). To find out how different estimations of the per gram cost affect the outcomes, Figures 3.8, 3.9, and 3.10 depict equilibrium consumption, $Q_{\hat{E}_2}$, and treatment's share of the budget at the steady state, $\hat{u}/gQ_{\hat{E}_2}$, for per gram costs at \$ 67, \$ 113 (base value), and \$ 200 as a function of the budget proportionality constant, g.

[12] Note that treatment-parameters *c* and *d* as well as the prevention-parameters *h*, *k* and *m* are linked and will, therefore, not be regarded separately.

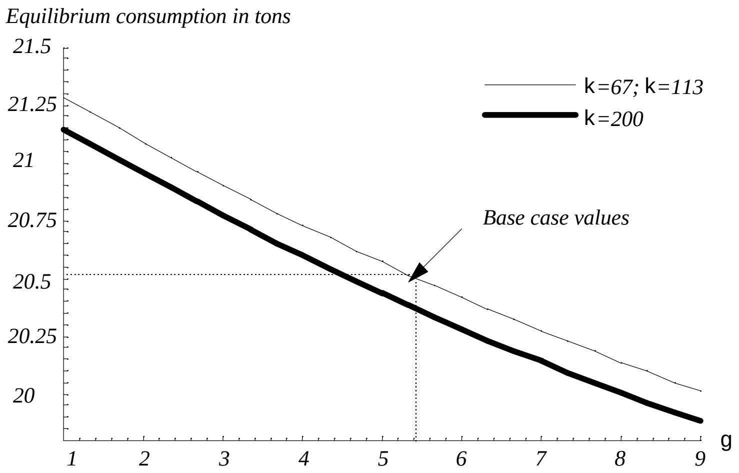

Figure 3.8

Equilibrium consumption as a function of the budget proportionality constant, γ, for social costs per gram κ = *67*, *113*, *200*

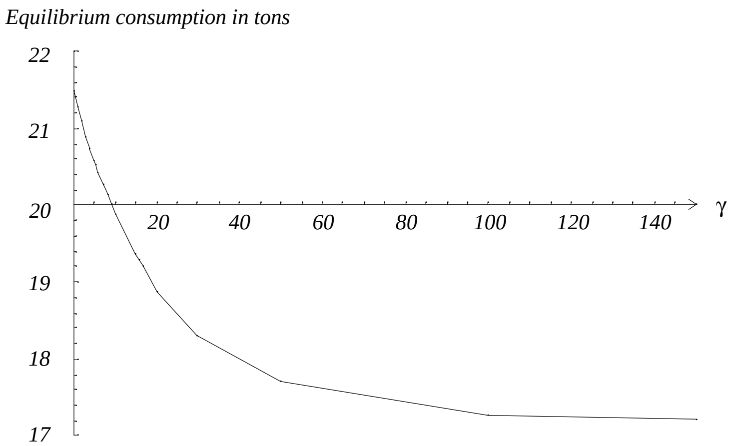

Figure 3.9

Equilibrium consumption, $Q_{\hat{E}_2}$, as a function of $\gamma \in [1,150]$; $\kappa = 113$

Figure 3.8 shows that a change in societal cost per gram has almost no effect on equilibrium consumption. In fact, if one estimates k to be about twice as large as its base value, equilibrium consumption decreases by 0.65%, while any underestimation of k has absolutely no effect. Whereas equilibrium consumption decreases when the government is willing to increase the budget per gram consumed. But even if the budget proportionality constant, g, amounts roughly 30 times the size of its base case value, equilibrium consumption decreases only by 3.3 tons, which is about 16%. Even for considerably high values of γ, it is not possible to eradicate illicit drug use. (See Figure 3.9.)

An interesting result follows from Figure 3.10. The higher the per gram costs, k and g, are estimated to be, the higher will be prevention's share of the equilibrium budget. One tries to decrease social cost, k$Q_{\hat{E}_2}$, by additional control measures, g$Q_{\hat{E}_2}$. For a rather high value of the social cost all resources go to prevention, because it's by far more cost-effective than treatment, e.g. $ 67 per student are sufficient to suppress 7.2% of initiation (Rydell, 1997).

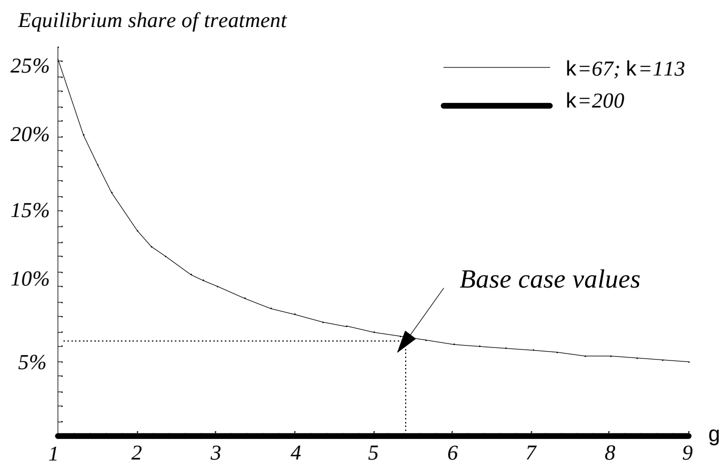

Figure 3.10

Steady state fraction of heavy users as a function of the budget proportionality constant, g, for social costs per gram k= *67, 113, 200*

Figure 3.11 illustrates the dependence of the values of the objective functional, J_2^*, and the discounted quantity totally consumed, both on different values of the social per gram cost, k, and of the budget per gram consumed, g.

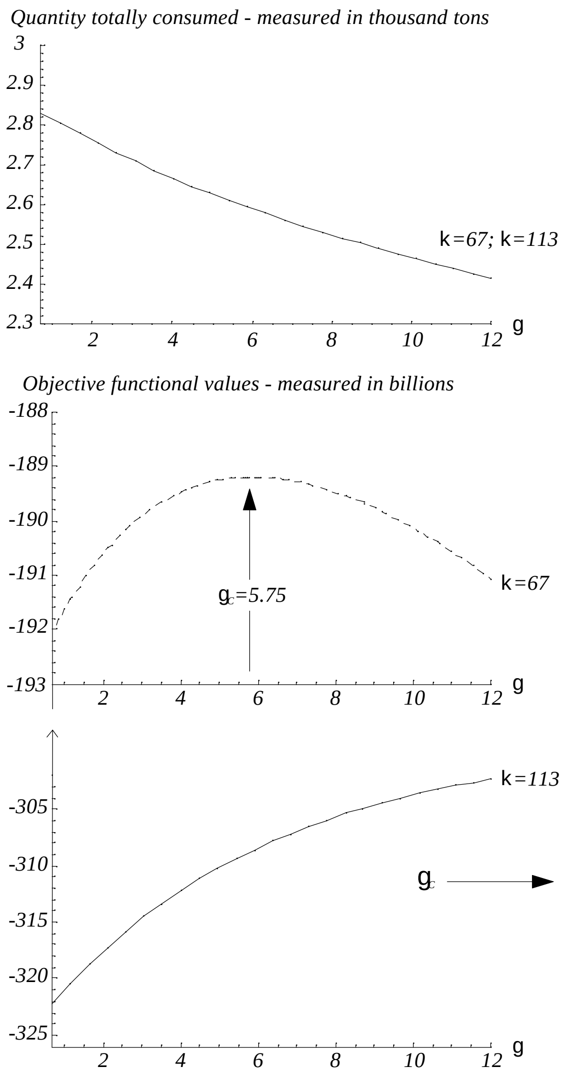

Figure 3.11

Discounted totally consumed quantity and objective functional value, J_2^*, as a function of g, for social costs per gram k=*67* and k=*113*; Initial data from 1996.

Again the quantity consumed is not affected by an underestimation of social cost per gram. The marginal reduction of the quantity consumed decreases with an increasing budget per gram. When the critical value, g_c, of the budget proportionality constant is reached, an additional dollar spent on control cannot be justified any more in the sense of minimizing the objective functional J.

The total financial loss, however, will be the bigger the higher the costs per gram, k, are, which makes sense. Secondly, it is a rather interesting feature that the slope of the optimal utility functional value curves as a function of g depends also on the size of k in a way described below. The peak in the utility functional curve as a function of per gram budget at g_c shifts forward for lower values of k and is more distinct. If the social costs are low ($ 67; dashed line), the government is better off with a small budget per gram consumed, whereas for higher per gram costs ($ 113; straight line) it would pay to increase the budget. But since the utility curves do exhibit maximum values, exceeding the critical value of g results much higher costs than advisable.

Since model (3.8) is parameterized with data from the current cocaine epidemic, Figure 3.11 suggests following: Either the budget (with prevention and treatment spending coming from the same funding source) is underestimated, or decision-makers might have placed very low estimates of the social costs in recent years, e.g. excluding „pain and suffering" costs. The base case parameter value of the per gram budget, $g=5.31$, amounts about one quarter of its critical value, for which the costs were minimized. Therefore people who perceive drug use to be costly for society should favor greater drug control spending per gram actually consumed.

3.2.2 Optimal allocation of the drug control budget and its costs for a drug epidemic like the one currently observed in the US

Since the shape of the controlled „market“ depends on the level of the initial numbers of users, we want to investigate the optimal control policies and their effects on the levels of use, corresponding societal cost and, cost of control for different stages of a drug epidemic, such like the one currently observed in the US. Similar to the results of section 3.1, also for the allocation problem (3.8) early interventions would have shortened the epidemic and reduced its intenseness drastically (see Figure 3.12 and 3.17).

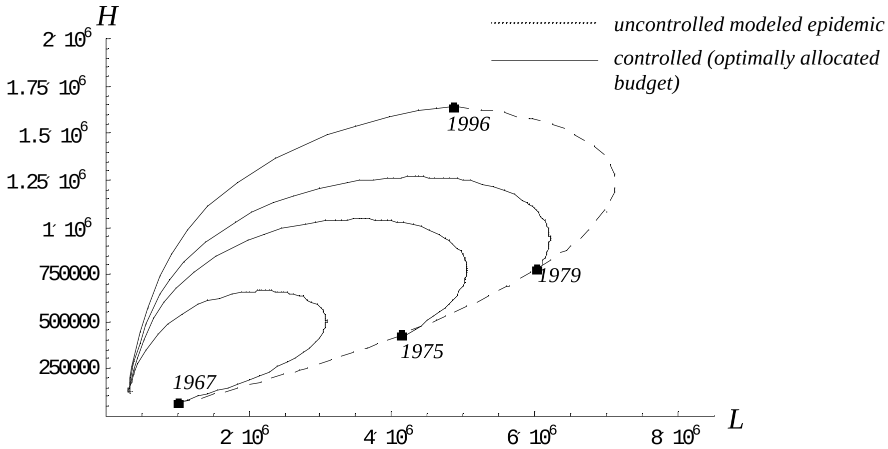

Figure 3.12

Projections of controlled trajectories (model (3.8)) into the *(L,H)*-plane - for initial years of the uncontrolled modeled epidemic (2.2)

A realistic scenario should take into consideration that it may take some time from the beginning of a drug epidemic until the government starts to impose controls on that specific drug problem. Many reasons could cause this lag: It takes a little while until people realize that a drug epidemic is going on, perhaps the government is not willing to intervene, or - even if the government is willing to do something - it may take some time until the bureaucracy is established, etc. Therefore, Figure 3.13 depicts the total discounted costs caused by the drug epidemic of concern as a function of *T*, where *T* denotes the time when government implements prevention and treatment, respectively. To put it in other words, we run the uncontrolled model (2.2) for *T* years which results in a specific number of addicts at time *T*, *L(T)* and *H(T)*. Starting with these initial conditions we run the optimally controlled allocation problem (3.8), with the standard infinite planning horizon assumption. Then, the absolute values of the optimal utility functionals from both problems are added by properly discounting the value of the second stage problem. As we can see from Figure 3.13, one could save up to 75% of the discounted resources by starting control early. Note that total discounted cost, *C(T)*, obtains a maximum at the peak in light use. Starting control measures later than that causes lower values of *C(T)* because one optimally supports the decline of the epidemic, instead of proactively driving the levels of use toward the equilibrium state.

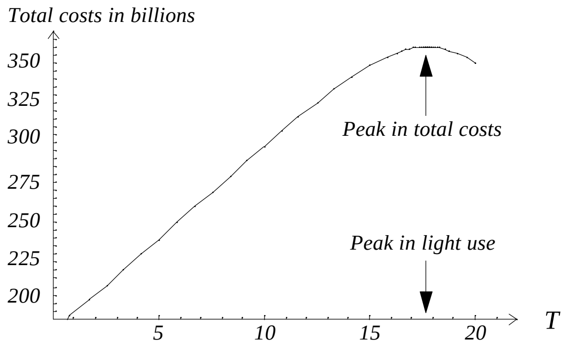

Figure 3.13

Totals costs, *C(T)*, as a function of *T*, where *T* denotes the time when government starts control; Calculation initiated with data observed in 1967.

Optimal policies for 1967

Starting with initial conditions that reflect US cocaine use in 1967, it would have been optimal (with a constrained budget) to put all drug control resources into prevention. No money should have been spent on treatment until 1981, at which point there would be a very rapid changeover from all prevention to all treatment. This transition would be completed in just six years. Treatment spending would then quickly rise a peak and then decay slowly as the amount of use fell. Then in 2013 prevention would begin to kick in again, taking over nearly all of the (much smaller) drug control budget approaching the equilibrium state. (See Figures 3.14, and 3.15.)

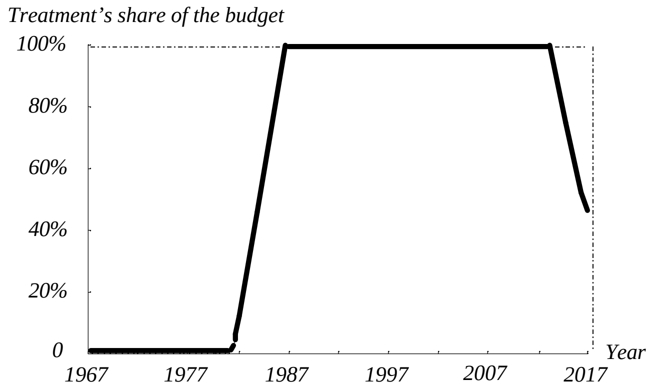

Figure 3.14

Treatment's share of the budget plotted versus time

The basic image here, depicted in Figure 3.15, is prevention first, then treatment, then back to mostly prevention as the epidemic approaches the steady state, with very abrupt changes from one regime to another. But how much would be spent on these interventions in order to reach the objective? I.e. how does the shape of the trajectories for control spending look like?

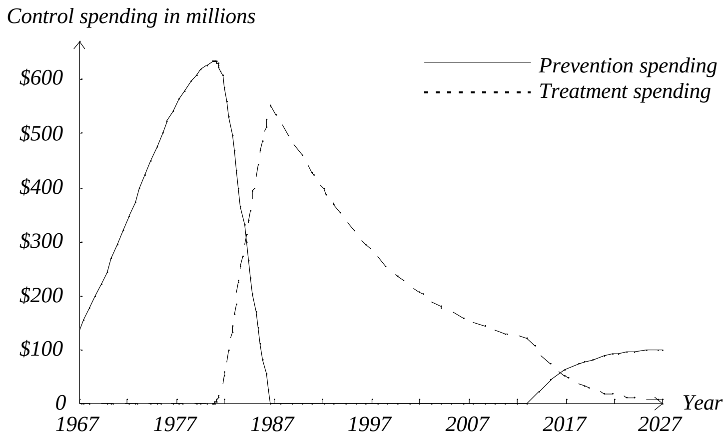

Figure 3.15

Optimal allocation of the drug control budget in millions

The sequence of peaks, with initiation peaking first, then light use, then the impact of prevention, then heavy use, then treatment is robust with respect to a wide range of parameter values, and the peaks are each separated by about three years. (See Figure 3.16.) Note that there is actually about a six-year lag between spending money on prevention and seeing its effect on initiation. Adding such a delay to the model might alter this sequence and their separation in time.

1975 →	1978 →	1981 →	1984 →	1987
↓	↓	↓	↓	↓
incidence	*lightuse*	*prevention*	*heavyuse*	*treatment*

Figure 3.16

Succession of peaks for initial data from the beginning of the cocaine epidemic

Due to an optimally allocated drug control budget light use would peak in 1978, at a level of 3.1 million (see Figure 3.17). This is less then 40% of the observed value. The number of heavy users would peak 17 years after starting optimal control measures, in 1984. After a maximum value of 0.6 million heavy users the numbers would have declined approaching the stable state which would have been obtained approximately 75 years after one decided to implement optimally allocated prevention and treatment programs.

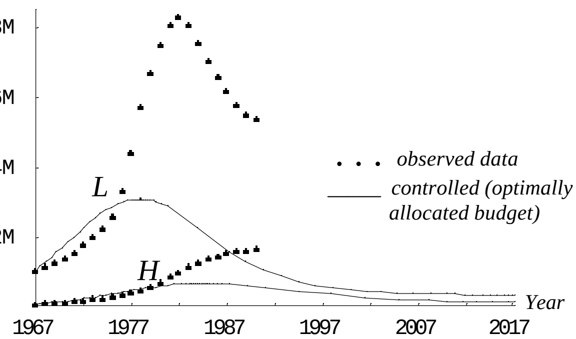

Figure 3.17

Time paths of controlled cocaine epidemic and smoothed historical data (Everingham&Rydell); Start of optimal budget allocation in 1967

For optimal control model (3.8) the fraction of heavy users, as an indicator for the severity of the drug problem, would peak in 1998 at a value of roughly 40%, whereas it would decrease ever after to its equilibrium value of 26% (see Figure 3.18). The high level of the equilibrium fraction explains why the equilibrium share of treatment is negligible. Since prevention emerges as the most cost efficient program - and all resources have to be spent even when they are not really needed - the preference would definitely not go to „expensive“ treatment. We argue that a problem without limitations on the financial resources would obtain a zero steady state for preventive measures.

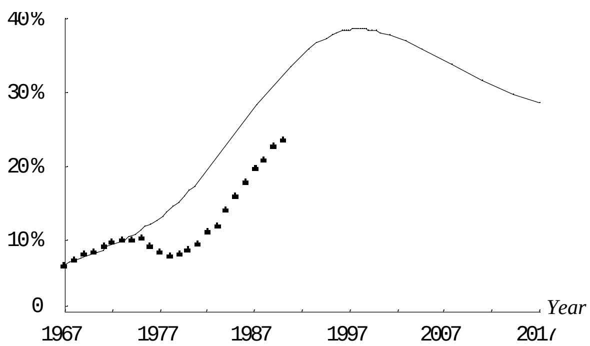

Figure 3.18

Time paths for fraction of heavy users of controlled cocaine epidemic and smoothed historical data (Everingham&Rydell); Start of optimal budget allocation in 1967

Figures 3.14 - 3.18 show what should have been done (subject to a constrained budget) starting in 1967, which is intellectually interesting but irrelevant since it is not 1967 anymore. What should we do moving forward from now?

Optimal control budget allocation from now on

Since the numbers of light and heavy users have not been changing so quickly in recent years, solving model (3.8) for what should have done starting with data observed in 1996 is of interest. The best strategy (subject to a constrained budget) suggests that we should have spent a few dollars on prevention in 1997 and 1998 (again, assuming no lag between spending and effect), but from 1999 through 2037, we should allocate all drug control resources to treatment. From

2037 to 2048 or so we should shift most of the (much smaller) resources back into prevention. (see Figure 3.19 and 3.20)

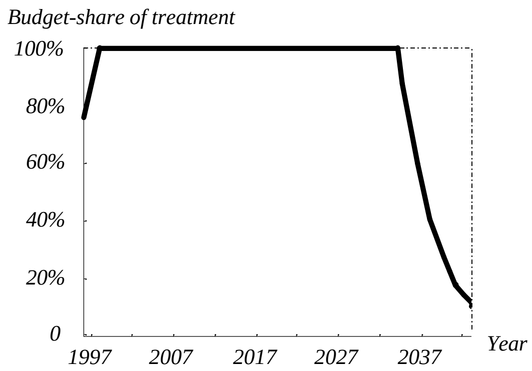

Figure 3.19

Treatment’s share of the budget plotted versus time

Since incidence and the number of light users peaked already in the 1980s treatment obtains its maximum value in 1998 and then slopes downward along the budget constraint until the year 2034 (see Figure 3.20). Since we „have to“ spend all the budget, ever after 2034 prevention spending decreases to obtain 93.3% of the equilibrium budget.

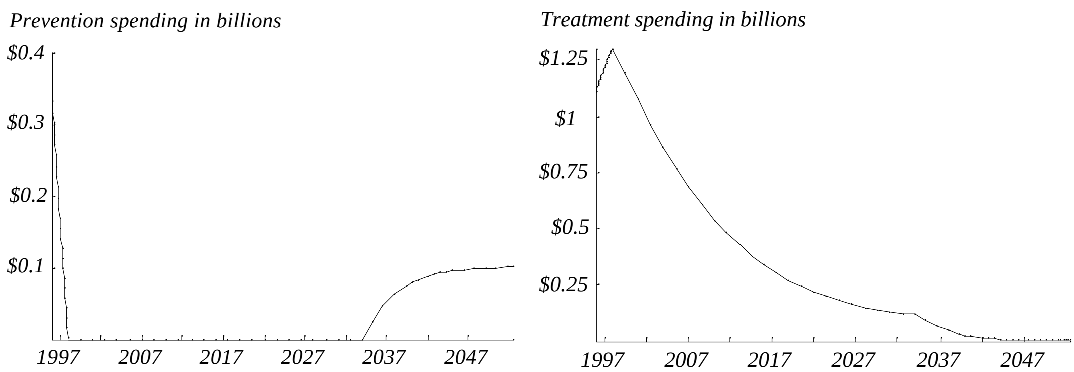

Figure 3.20

Optimal evolution of prevention and treatment spending, respectively

The basic pattern for the 1967 data and those observed in 1996 is the same with two exceptions. First, we are already close to the point when prevention should not receive any of the scarce dollars. Second, the period during which we should rely entirely on treatment is longer starting from 1996 conditions than it would be if we would started this control in 1967.

Additionally, to wait out so long until implementation of the optimal allocation of resources has its price: On the one hand the „1967-plan“ would reduce cost to 63% of the present value, on the other hand the fraction of heavy users maintains to increase and reaches the 50% mark of prevalence in 2012. While the majority of the light users remains employed, finances their habit with legitimate income and avoids severe, adverse health consequences, heavy users are less fortunate. Therefore very high societal cost is associated with them.

3.3 Optimal control without financial limitations on the interventions

In this section we want to derive optimal policies for prevention spending, w, and expenditures on treatment, u, without any budget restriction. At some stages of the epidemic expenses higher than a specific proportion of the quantity consumed, $Q=16.42L+118.93H$, might be useful, at other stages less money was sufficient. Additionally this model is more appropriate than one with a constrained budget if treatment and prevention resources are not allocated from a single pot.

Again the government's objective is to minimize the costs to society, which are proportional to the quantity of nation-wide consumed cocaine, as well as the costs of control. We consider control problem (3.11) with two states, L and H, and two controls, w and u.

$$w^* = \arg\max_{0 \le w} J_3$$

$$u^* = \arg\max_{0 \le u} J_3$$

with

$$J_3 = -\int_0^\infty e^{-rt}\left(\kappa\, Q + w + u\right) dt \qquad (3.11)$$

subject to $\quad \dot{L} = I(L,H)\psi(w) - (a+b)L, \qquad L(0) = L_0$

$\qquad\qquad\quad \dot{H} = bL - (g + \beta(H,u))H, \qquad H(0) = H_0$

Since we expect prevention to extinct as the equilibrium state is reached the analysis of model (3.11) proceeds in a slightly different manner compared to the optimal allocation model (3.8). Following equations summarize the necessary conditions for the optimal policies w^* and u^*, all in current values, where $\mathcal{H}=\mathcal{H}(L,H,\pi_1,\pi_2,w,u)$ denotes the Hamiltonian, $\mathcal{L}=\mathcal{L}(L,H,\pi_1,\pi_2,w,u,\theta,\sigma)$ the Lagrangean and π_1 and π_2 the costate variables:

$$\mathcal{H} = -\kappa(16.42L + 118.93H) - u - w + \pi_1(I\psi - (a+b)L) + \pi_2(bL - (g+\beta)H) \tag{3.12a}$$

$$\mathcal{L} = \mathcal{H}(L,H,\pi_1,\pi_2,w,u) + \mu w + \nu u \tag{3.12b}$$

$$w^* = \arg\max_{w} \mathcal{L} \tag{3.13a}$$

$$u^* = \arg\max_{u} \mathcal{L} \tag{3.13b}$$

Immediately from the Hamiltonian maximizing conditions (3.13) one derives that both costates are negative and that both controls are truly maxima. (See Appendix A3.3.) The derivation of the canonical system for the unconstrained problem (3.11) is given in Appendix A3.4. Note that the only restriction is the standard non-negativity assumption for control spending, which has to be met in all models presented here. Particularly we have to distinguish between optimal interior policies for the expenses on prevention and treatment and optimal interventions at the border of the admissible region, i.e. $u^*=0$ and or $w^*=0$.

3.3.1 Equilibrium consumption and parameter sensitivity

Equipped with the base case parameter values, listed in Table 3.1, we compute the equilibrium values of states, costates, and controls by simultaneously solving $\dot{L}=0$, $\dot{H}=0$, $\dot{\pi}_1=0$, and $\dot{\pi}_2=0$ for the canonical systems given in Appendix A.3.4. Only canonical system (A3.15) possesses an feasible steady state, i.e. we obtain an equilibrium at the border of the admissible region ($w^*, u^* \geq 0$) where prevention spending is equal to zero. As was the case for the optimal allocation problem, each pair of initial values defines a unique trajectory on the stable manifold, which is the optimal solution of system (3.11).

$$\hat{E}_3 = \begin{pmatrix} \hat{L} \\ \hat{H} \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{w}^* \\ \hat{u}^* \end{pmatrix} = \begin{pmatrix} 356{,}958 \\ 130{,}997 \\ -36{,}619.4 \\ -14{,}074.6 \\ 0 \\ 4{,}779{,}340 \end{pmatrix}$$

Compared to the uncontrolled epidemic (model (2.2)) equilibrium consumption according to model (3.11) remains approximately equal in size, and the equilibrium fraction of heavy users, $\hat{H}/(\hat{L}+\hat{H})$, decreases slightly. (See Table 3.4.) Though treatment spending constitutes the entire equilibrium budget, the resources spent on it are negligible.

Table 3.4: Equilibrium quantities of model (3.11)

$Q_{\hat{E}_2}$	$\frac{\hat{H}}{\hat{L}+\hat{H}}$	$\frac{\hat{u}^*}{\hat{w}^*+\hat{u}^*}$
21,444,100 grams	26.8 %	100 %

Table 3.5 summarizes the effects of 1% changes from the base case parameter values (Table 3.1) on the steady state, equilibrium consumption, the fraction of treatment of the budget, as well as the fraction of heavy users. As it was the case for the optimal allocation model, t is proportional to the equilibrium states and equilibrium consumption, while the fraction of heavy users is not affected at all by a change in the number of innovators. Apart from this, the numbers of light and heavy users and equilibrium consumption react mainly to changes in the rate at which light users quit, *a*, and the measure of deterrence, *q*. Parameters *h* and k do not affect the equilibrium quantities, and equilibrium prevention as well as its share of the budget is totally insensitive to any of the parameter changes.

The marginal efficiency of treatment, *d*, affects nothing else but size of equilibrium treatment. If one was able to improve the effect of an extra dollar spent on treatment, less resources would be needed to achieve the same result and, consequently, the unconstrained control policy would only demand control spending in the very beginning of the epidemic.

Table 3.5: Effects of a 1% shortage of the base case parameter values on the equilibrium state $\hat{E}_3$, equilibrium consumption $Q_{\hat{E}_3}$, fraction of heavy users $\hat{H}/(\hat{L}+\hat{H})$, and treatment's share of the budget.

-1%	$\hat{L}$	$\hat{H}$	$\frac{\hat{H}}{\hat{L}+\hat{H}}$	$Q_{\hat{E}_3}$	$\hat{u}^*$	$\hat{w}^*$	$\frac{\hat{u}^*}{\hat{w}^*+\hat{u}^*}$
a	+ 0.95%	+ 1.21%	+ 0.19%	+ 1.14%	0%	0%	0%
b	+ 0.39%	+ 0.14%	- 0.18%	+ 0.21%	0%	0%	0%
g	- 0.21%	+ 0.03%	+ 0.18%	- 0.03%	0%	0%	0%
t	- 1%	- 1%	0%	- 1%	0%	0%	0%
s	- 0.03%	- 0.39%	- 0.26%	- 0.29%	0%	0%	0%
q	- 0.32%	+ 0.96%	+ 0.47%	+ 0.78%	0%	0%	0%
d	- 0.06%	- 0.01%	- 0.05%	+ 0.01%	+ 3.25%	0%	0%
h	0%	0%	0%	0%	0%	0%	0%
k	- 0.02%	0%	+ 0.02%	0%	0%	0%	0%
r	- 0.05%	+ 0.01%	+ 0.04%	- 0.01%	0%	0%	0%

3.3.2 Optimal drug control spending

The general pattern of drug control spending looks similar to the optimal allocation of resources discussed in the last subsection apart from their actual size, however. The total budget spent in the unrestricted case is significantly higher early in the epidemic. While the constrained budget is limited to grow with the quantity consumed, unrestricted prevention amounts a multiple of the optimally allocated value right from the beginning of the epidemic if the size is free. In fact, both treatment and prevention spending amount a multiple of the optimally allocated values during the first 30-40 years of the epidemic. Ever after practically no resources are needed for prevention or treatment programs. (See Figure 3.21 and 3.22.)

Starting with initial conditions that reflect US cocaine use in 1967, it would be optimal to focus entirely on prevention. No money should be spend on treatment until 1980, at which point there would be a very rapid changeover from all prevention to all treatment spending. This transition would be completed in just six years. Treatment spending would peak immediately after and then decay towards its equilibrium value.

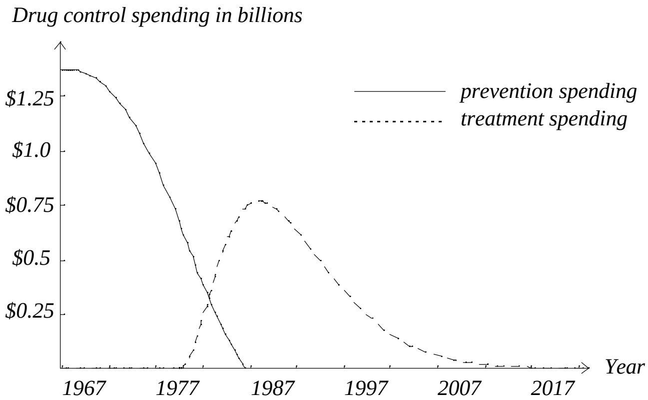

Figure 3.21

Prevention and treatment spending for the unrestricted case; Initial data from 1967

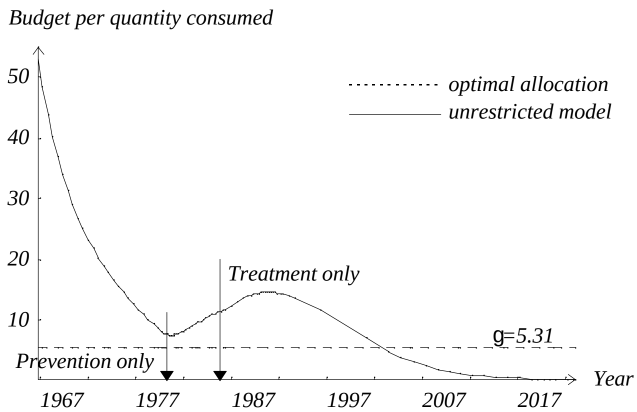

Figure 3.22

Budget per gram consumed for optimally allocated and unconstrained model; Initial data from 1967

Figure 2.22 reflects the budget per gram consumed. Contrary to the optimal allocation problem for the unrestricted model this value varies over time. As we see the slopes of the “budget per gram“-curves intersect only once. In the first 35 years of the epidemic the budget demand is always more than $g=5.31$. Ever after practically no financial resources are necessary.

Even though lots of additional resources are demanded in the beginning of the epidemic, the value of the objective functional is cut back to 27% of its uncontrolled value and to 55% of the optimally allocated value. This is remarkable because the choice of the time preference rate stresses out that we are more interested in what happens in the nearer future. For the unrestricted control policy total costs are smaller at any time *T* compared to the optimal allocation control policy, even though total costs come very close in the intermediate years of the epidemic. (See Figure 3.23.) But nevertheless it might be difficult for decision-makers to justify these high expenditures in the beginning of an epidemic. When the drug problem is perceived to be small, the public might resist having many tax dollars being spent on drug control; they might prefer that scarce resources be allocated to more pressing issues. On the other hand, when the drug problem is perceived to be severe, the public might demand that significant interventions are made.

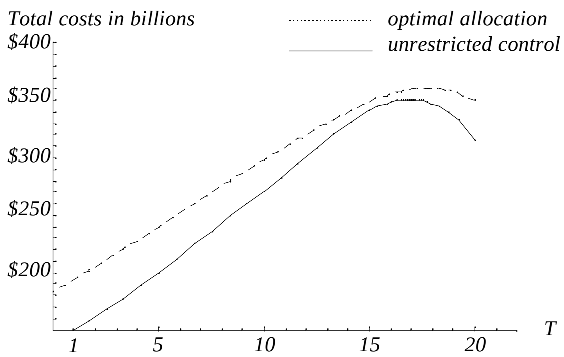

Figure 3.23

Total costs for optimal allocation and unrestricted control as a function of *T*, where *T* denotes the delay in the starting year of control; Calculation started for 1967

Chapter 4

Comparison of control policies

4.1 Equilibria and objective functional values

In this chapter we will show up the differences in efficiency and cost of the constant fraction budget rule (section 3.1), another budget policy where one is free to alter the proportion of allocation between prevention and treatment (section 3.2), and the unconstrained problem (section 3.3). Very different to the transients the state equilibrium quantities of the models considered here just slightly differ from each other (Table 4.1). To find out more details on this, Figure 4.1 illustrates the percent values of equilibrium consumption relative to the (100%) case of no control.

Table 4.1: Equilibrium quantities of state and control

Control	$\hat{L}$	$\hat{H}$	$\hat{w}*$	$\hat{u}*$
none	338,416	132,864	0	0
constant fraction, $f^*=0$	323,055	126,834	108,265,222	0
optimal budget allocation	347,037	124,646	101,631,000	7,344,250
unconstrained problem	356,958	130,997	0	4,779,340

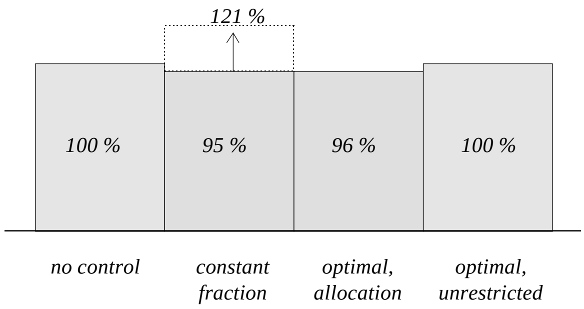

Figure 4.1

Percent values of equilibrium consumption for all models relative to the uncontrolled value

Note that for the constant fraction model, Figure 4.1 shows the entire range of possible values of equilibrium consumption, contrary to Table 4.1 which only gives a description of the special case $f^*=0$. The quantity consumed at the steady state may increase up to 121% of its uncontrolled value, if the epidemic is in an initial state which favors large budget-shares for treatment. Then one is obligated to pursue a given level of treatment throughout the course of the epidemic. Treating and removing heavy users when there are relatively few heavy users is counter-productive because each user contributes significantly to deterring initiation. If there are many heavy users relative to the number of light users, however, treating and removing some does not greatly jeopardize the negative reputation of the drug or the suppression of initiation. Though a constant fraction policy might have its advantages, e.g. low implementation costs, its application to a dynamic process which claims for intertemporal controls is not advisable. Especially due to the cyclical behavior for data observed at the present stage of the US cocaine epidemic the constant fraction budget rule definitely needs improvement.

This leaves us with the optimal dynamic control policies. Since a neighborhood of the equilibrium state is reached in 100 years or so which is „rather uninteresting“ due to the choice of the time preference rate, the equilibrium values are very similar. What really differs are the ways to get there and the associated costs. Since these costs depend obviously on the initial state of the epidemic one should investigate at least several initial conditions specific for a particular stage of a drug epidemic like the one currently observed in the US. (See Table 4.2 and Figure 4.2.)

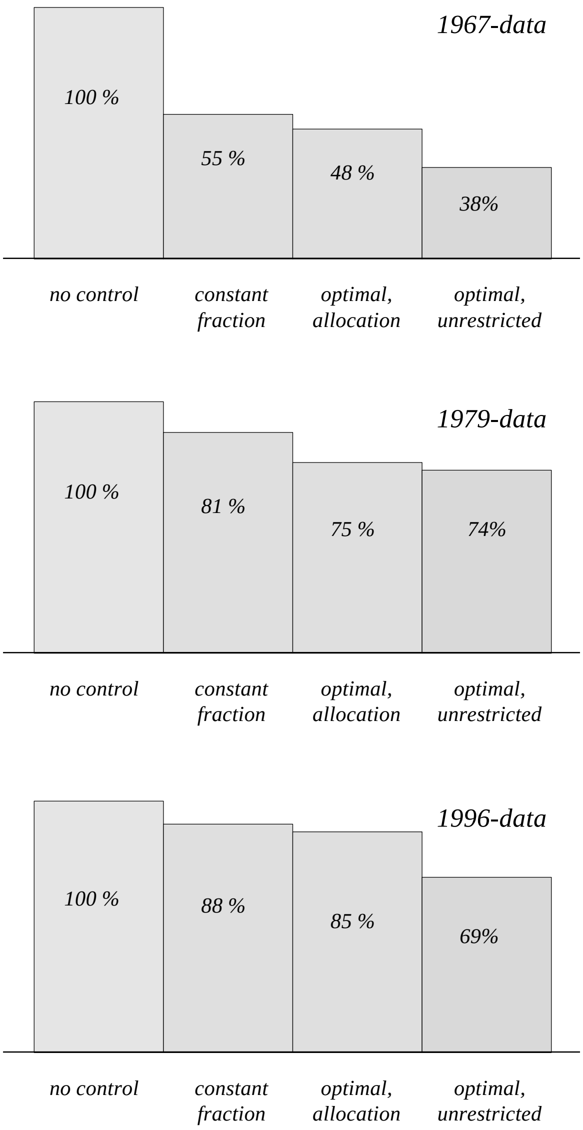

Figure 4.2

Percent values of the objective functionals, J, for constant fraction, optimal allocation and unrestricted control problems relative to the uncontrolled value

Table 4.2: Comparison of the values of the objective functional *J** in millions

Control	J*(1967)	J*(1975)	J*(1979)	J*(1996)
none	- 399,744	- 484,295	- 492,902	- 364,586
constant fraction	- 218,393	- 341,661	- 401,474	- 321,010
optimal allocation	- 193,735	- 318,445	- 367,885	- 309,411
unrestricted	- 150,389	- 290,076	- 365,483	- 250,084

Implemented early in the epidemic the two allocation problems can cut back total social loss remarkably. The unrestricted control, however, by far reduces the total loss most efficiently. Also for data observed at the present stage of the epidemic the value of the objective functional shrinks by some 30% due to unrestricted control. For intermediate cases, however, there is almost no difference between optimal allocation and unrestricted control. The quantity consumed is already so huge that the size of the budget, defined to be proportional to it, is similar to unrestricted control spending. Whereas, in the beginning and at later stages of the epidemic the difference between these two „budgets“ is striking.

4.2 Evolution of control spending and levels of use

Figure 4.3 shows the projections of the optimally controlled trajectories into the *(u,w)*-plane. Unrestricted control requests a multiple of constrained budget prevention spending in the beginning of the epidemic (in 1967) and $0.2 billion more treatment spending at later stages of the epidemic. The budget share of treatment, however, looks very similar for optimal allocation and unrestricted control with a transition from all prevention to all treatment spending happening with the same rate (in approximately six years). Particularly, it is obvious from Figures 4.3 and 4.4 that no prevention spending is needed approaching the steady state in the framework of the unrestricted control problem. But since we did not build in a delay between prevention spending and seeing its effects a certain level of preventive measures might be an appropriate mean to establish awareness of the danger of drug consumption in people’s minds to avoid possible future drug epidemics. Full scale prevention spending according to the constant fraction model gives a totally different picture and is not able to reduce the total social loss like the other two policies do.

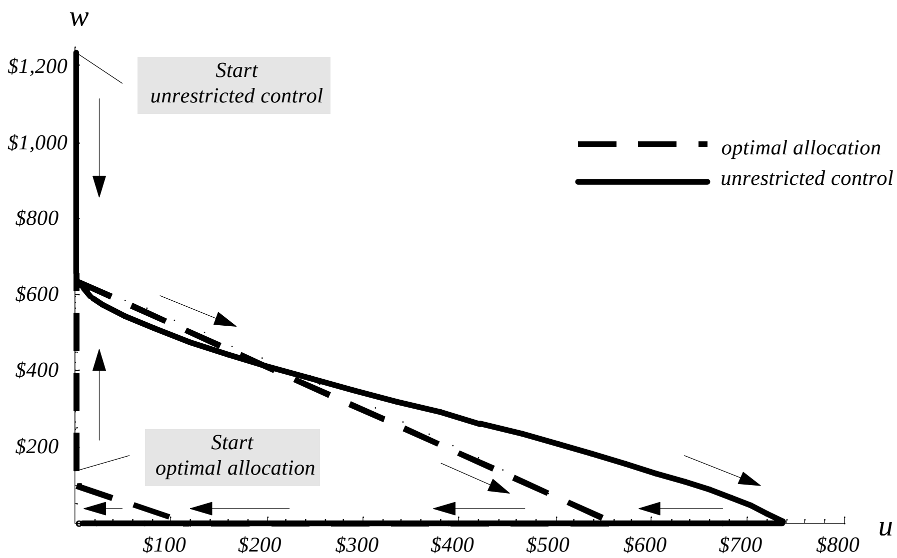

Figure 4.3

Projections of controlled trajectories into the *(u,w)*-plane depicting treatment and prevention spending in million dollars; Initial data from 1967

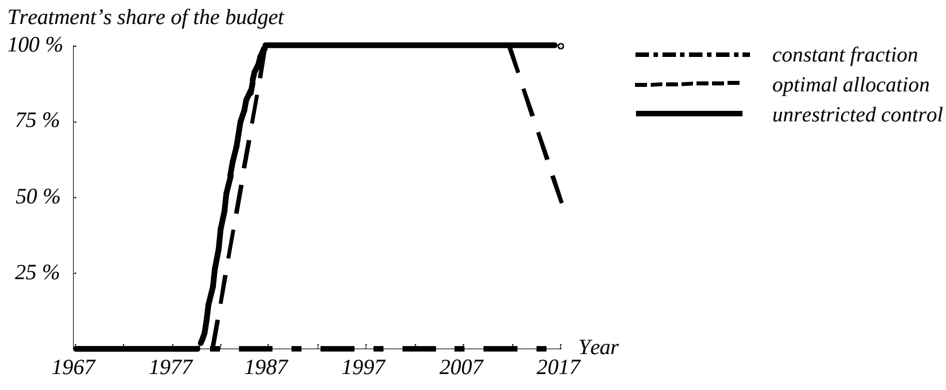

Figure 4.4

Time path of the proportion of the control budget going to treatment for constant fraction, optimal allocation and unrestricted control; Initial data from 1967

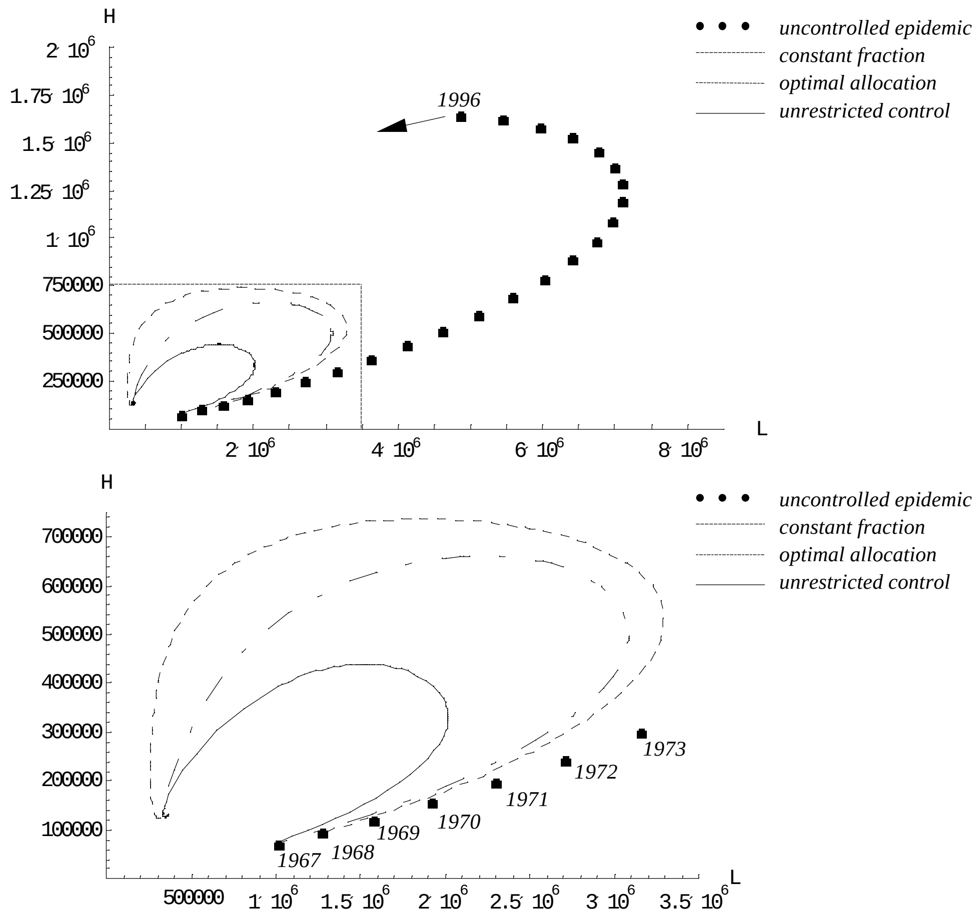

Figure 4.5

Projections of controlled trajectories into the *(L,H)*-plane - for initial years of the uncontrolled modeled epidemic (model (2.2)); Initial data from 1967

Figure 4.5 shows the respective path of the uncontrolled trajectory as well as the projections of the controlled trajectories into the *(L,H)*-plane. Qualitatively these curves look very similar. Controlled data move around their respective path and approach equilibria which fall close together. Apart from the unrestricted control, the trajectories stay close together during the first years, especially those of the two allocation problems. These trajectories diverge as soon as any share for treatment programs is requested, since the optimal budget allocation may vary over time in model (3.8). For initial data from later stages of the epidemic the divergence is clearly distinct, because constant fraction treatment measures have to be implemented „too early“ in order to have them „later“, when really needed, and intensify the problem. For initial data from intermediate phases of the epidemic, where consumption and the associated drug control budget is very high, the trajectories of the optimal allocation and the unrestricted control move close together, clearly separated from the course of the sub-optimal constant fraction control.

4.3 Decision-making support

Figure 4.6 informs of the following decision. If one has to decide today and for all time what proportion of a constrained budget control will go to prevention and what proportion will go to treatment, then as a function of the current state, this is what do to. Since this is a pretty artificial problem, just note that for most initial states, one will want to do all prevention or all treatment. Particularly one will want to do all prevention if there are a lot of light users relative to heavy users and vice versa.

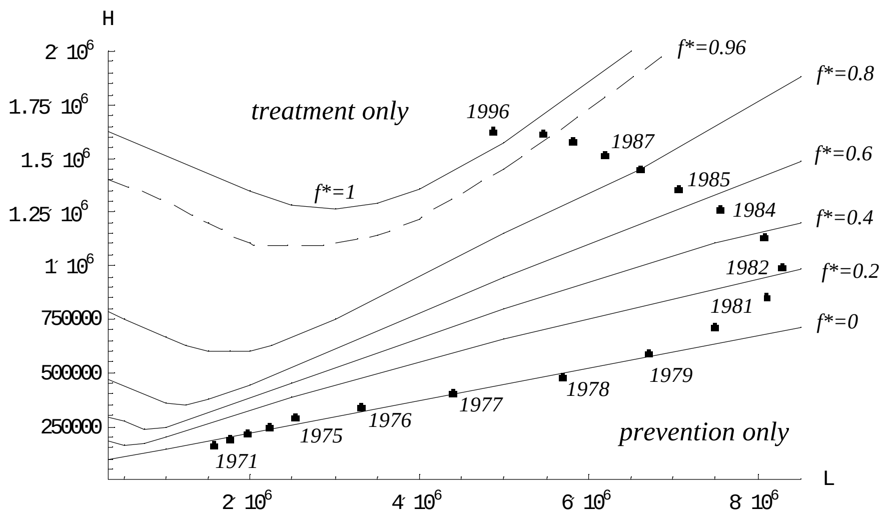

Figure 4.6

Policy isocline chart for the fixed proportion, constrained budget problem

Figure 4.7 is the comparable policy guide for the constrained budget, optimal allocation problem. It informs, as a function of the current state of the epidemic, how one should divide the drug control budget between treatment and prevention at this particular point in time, assuming that treatment and prevention come from the same fixed funding source. Given a current estimate of the number of light and heavy users, one can read of f_0. Over time, the number of light users will change, and that optimal f_0 will change too, but for any given starting point that trajectory can be added to this chart.

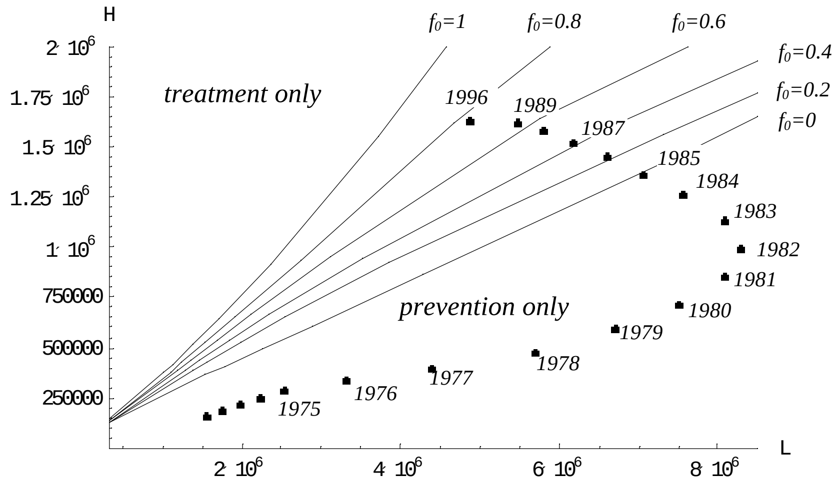

Figure 4.7

Policy isocline chart for the constrained budget, optimal allocation problem

The range where the allocation of the budget does not entirely go to prevention or treatment looks different for the two models with a constrained budget. Since the proportion of allocation is allowed to vary over time in model (3.8) the range shifts further north and shrinks. Actually the all prevention/all treatment isoclines are almost identical with the $\dot{L} = 0 / \dot{H} = 0$ isoclines for optimal allocation control. In other words, without a delay between prevention spending and seeing its effects one should stop preventive measures immediately after the peak in light use is observed and invest all resources in treatment after the peak in heavy use has happened to support the decline of the epidemic optimally (with a constrained budget). In the transition years from the peak in light use to the peak in heavy use, which takes about six years for trajectories crossing the *(L,H)*-plane from south-east to north-west reasonably close to the equilibrium, the resources according to the optimally allocated control budget will be spent on both interventions. (See Figure 4.7.)

For the unrestricted control problem the projection of the optimal budget allocation into the *(L,H)*-plane, depicting the policy isoclines, looks similar to Figure 4.7 apart from an even smaller transition range from all prevention to 60% treatment and a wider transition range from 20% prevention to all treatment for trajectories crossing the *(L,H)*-plane from south-east to north-west far away from the equilibrium. (See Figure 4.8.) E.g. following the course of the observed cocaine epidemic the transition from all prevention to all treatment takes about nine years (all prevention $\xrightarrow{\text{2 years}}$ 60% treatment $\xrightarrow{\text{2 years}}$ 80% treatment $\xrightarrow{\text{5 years}}$ all treatment). For trajectories crossing the *(L,H)*-plane reasonably close to the equilibrium, e.g. starting control in 1967, the transition happens rather similar to the optimal allocation problem (see also Figure 4.4).

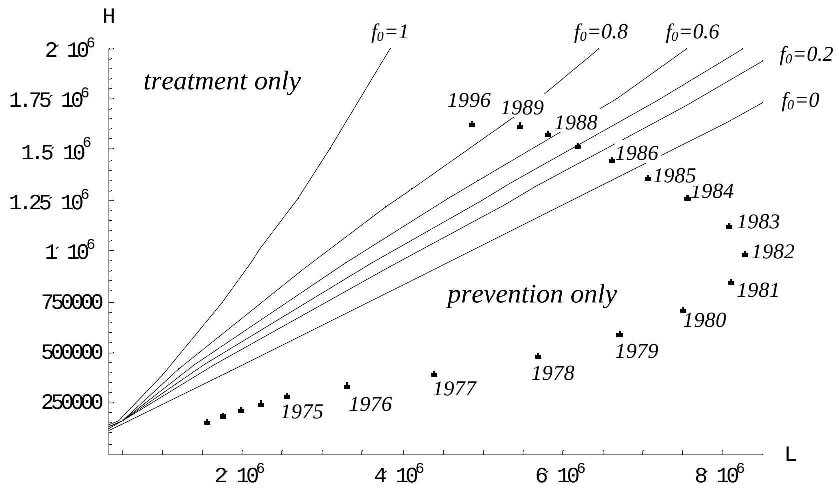

Figure 4.8

Policy isocline chart for the unrestricted optimal control problem

Since control spending in the unrestricted case is not proportional to the quantity totally consumed any more, Figure 4.9 depicts the prevention and treatment spending, respectively, in million dollars for each combination of light and heavy users.

Chapter 5

Conclusion

The models presented here generate a number of interesting insights, among which are

(1) Control measures, such as treatment and prevention, are most appropriate for specific stages of a drug epidemic and should change over time. Static interventions applied to a dynamic process may be counter-productive. Particularly prevention does best when there are relatively few heavy users, i.e. in the beginning of an epidemic. Treatment is very sufficient to support the decline of drug abuse optimally.

(2) Some control, even a „dumb“ one as constant fraction, does better than no control. Optimal unconstrained control does much better than constrained budget control does. This is even true though differences in equilibrium are not great; the differences show up during the transition to steady state.

(3) One should try to quickly detect the onset of an epidemic and start optimal control measures early to lower the total social loss. Both, costs of interventions as well as the social cost associated with the quantity consumed increase with the delays in the starting year of control.

(4) People who perceive drug use to be costly for society should favor greater drug control spending per gram consumed and allocate a grater 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, even if it might be difficult to justify this behavior by the „magnitude of the problem“.

Appendix to Chapter 2

A.2.1 Local Stability Analysis

Since the initial conditions are always positive, L_0, $H_0>0$, and the first quadrant is an invariant set we restrict our analysis to the positive quadrant of the *(L,H)*-plane. First we derive the equations for the isoclines:

$$\dot{H} = 0: \quad H(L) = \frac{b}{g}L \tag{A2.1}$$

$$\dot{L} = 0: \quad H(L) = -\frac{L}{q}\ln\left[\frac{a+b}{s} - \frac{t}{sL}\right], \text{ defined for } \frac{t}{a+b} < L \tag{A2.2}$$

The intersection of the isoclines uniquely defines an equilibrium $\hat{E}$ inside a sub-section of the positive quadrant of the *(L,H)*-plane $(t/(a+b) < L,\ 0 < H)$ if and only if $t > 0$ and $\Omega := a + b - s\exp\left[-q\frac{b}{g}\right] > 0$ *:*

$$\hat{E} = \begin{pmatrix} \hat{L} \\ \hat{H} \end{pmatrix} = \frac{t}{\Omega}\begin{pmatrix} 1 \\ b/g \end{pmatrix} \tag{A2.3}$$

Equation (A2.3) also reflects the equilibrium ratio $L : H = 1 : b/g$, which leads to a fraction of heavy users of $b/(b+g)$. The local stability behavior of system (2.2) can be determined in the usual manner by the linearization around $\hat{E}$ using the Jacobian matrix evaluated at the stationary point:

$$\Im\big|_{\hat{E}} = \begin{pmatrix} \left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - \Omega & -\ sq\ \exp\left[-q\frac{b}{g}\right] \\ b & -g \end{pmatrix}$$

and the corresponding characteristic roots are calculated as follows:

$$\lambda_{1,2} = \frac{1}{2}\left\{\left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - g - \Omega \pm \sqrt{\left(\left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - g - \Omega\right)^2 - 4\gamma\Omega}\right\} \tag{A2.4}$$

Based upon these values (A2.4) we are able to characterize the local stability behavior of the equilibrium point (A2.3):

(i) The roots are real, distinct, and of the same sign, i.e. $\lambda_1\lambda_2 > 0$. The trajectories will be locally stable if the roots are negative and unstable if they are positive. The equilibrium has the properties of a *node*.

(ii) The roots are real, distinct and of opposite sign, say $\lambda_1 < 0 < \lambda_2$. The equilibrium is called a *saddle point*. The trajectories approach the equilibrium along the stable manifold and leave it via the unstable manifold.

(iii) The roots are conjugate complex, $\lambda_{1,2} = \rho \pm \omega i$, with nonzero real parts ($\rho \neq 0$). If $\rho < 0$ than the trajectories spiral towards the equilibrium; it is called *a stable focus*.
If $\rho < 0$ than the trajectories form a *unstable focus* spiraling away from the equilibrium.

(iv) If $\rho = 0$ than the equilibrium is either a *focus* or a *center*. In the latter case, instead of spiraling away or towards the equilibrium, the trajectories are closed circles. Note that for a center the equilibrium is stable but not asymptotically so.

A.2.2 The Hopf Bifurcation Theorem [13]

Suppose that a family of time-continuos dynamical systems $\dot{z} = F(z, \mu)$, F continuously differentiable, $z \in \Re^n$ and $\mu \in \Re$, has an equilibrium $\hat{z}_c$ at the value μ_c of the parameter, at which the following properties are satisfied:

[13] There exist several versions of the Hopf bifurcation theorem. The following one is taken from Guckenheimer and Holmes (1983), p. 151ff,. For other versions see, e.g. Alexander and Yorke (1978) and Marsden and McCracken (1976)

(i) The Jacobian evaluated at $\left(\hat{z}_c, \mu_c\right)$,

$$\Im(\mu_c) := \left.\frac{\partial F(z,\mu)}{\partial z}\right|_{\mu=\mu_c,\, z=\hat{z}_c}, \tag{A2.5}$$

has eigenvalues $\lambda(\mu)$ that vary smoothly with μ and has at $\mu = \mu_c$ a simple pair of purely imaginary eigenvalues, say $\lambda_{1,2} = \pm\omega i$, and no other eigenvalues with real parts zero; μ_c is called a critical value of the bifurcation parameter μ or, for short, the bifurcation point. Then there exists a smooth curve of equilibria depending on the bifurcation parameter μ, denoted $\hat{z}(\mu)$ so that $\hat{z}(\mu_c) = \hat{z}_c$.

(ii) If, moreover,

$$\left.\frac{d\left(\mathrm{Re}\,\lambda(\mu)\right)}{d\mu}\right|_{\mu=\mu_c} = D \neq 0 \tag{A2.6}$$

then there is a unique three-dimensional center manifold passing through $\left(\hat{z}_c, \mu_c\right)$ in $\Re^n \times \Re$ and a smooth system of coordinates (preserving the planes $\mu = const$) for which the Taylor expansion of degree 3 on the center manifold is given by the following normal form

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} D(\mu-\mu_c) & -\left(\omega + C(\mu-\mu_c)\right) \\ \omega + C(\mu-\mu_c) & D(\mu-\mu_c) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} x\left(x^2+y^2\right) \\ y\left(x^2+y^2\right) \end{pmatrix}. \tag{A2.7}$$

If $A \neq 0$, there is a surface of periodic solutions in the center manifold which has quadratic tangency with the eigenspace of $\lambda_{1,2}(\mu_c)$ agreeing to second order with the paraboloid $\mu = -\frac{A}{D}\left(x^2+y^2\right)$.

(iii) If $A<0$, then these periodic solutions are stable limit cycles, while if $A>0$, the periodic solutions are repelling.

A.2.3 Evidence of Supercritical Hopf Bifurcation for any of the System-Parameters ν (ν = *a*, *s*, *q*, *b*, *g*)

At any Hopf bifurcation point $\left(\hat{E}_c, \nu_c\right)$ ($\hat{E}_c$ is determined by Equation (A2.3)) the normal form of model (2.2) is given by following system, where $\lambda_{1,2}(\nu_c) = \pm i\omega_c$ denotes the corresponding pair of purely imaginary eigenvalues[14]

$$\begin{pmatrix} \dot{L} \\ \dot{H} \end{pmatrix} = \begin{pmatrix} 0 & -\omega_c \\ \omega_c & 0 \end{pmatrix} \begin{pmatrix} L - \hat{L}_c \\ H - \hat{H}_c \end{pmatrix} + \begin{pmatrix} \varphi(L,H) \\ \psi(L,H) \end{pmatrix}, \text{ where } \begin{pmatrix} \varphi(L,H) \\ \psi(L,H) \end{pmatrix} = \begin{pmatrix} I(L,H) - (a+b)L + \omega_c\left(H - \hat{H}_c\right) \\ bL - gH - \omega_c\left(L - \hat{L}_c\right) \end{pmatrix}$$

with $\varphi\left(\hat{E}_c\right) = \psi\left(\hat{E}_c\right) = 0$ and $\frac{d}{dt}\varphi\left(\hat{E}_c\right) = \frac{d}{dt}\psi\left(\hat{E}_c\right) = 0$. As long as assumption (A2.8) holds at $\left(\hat{E}_c, \nu_c\right)$, where accordingly Conditions (A2.5) and (A2.6) are satisfied,

$$3g - qb > 0, \qquad \text{(A2.8)}$$

the normal form calculation yields a negative cubic coefficient A (for the two-dimensional system[15])

$$A = \frac{1}{16}\underbrace{\left[\underbrace{I_{LLL}}_{<0} + \underbrace{I_{LHH}}_{<0}\right]}_{<0} + \underbrace{\underbrace{\frac{1}{16\omega_g}}_{>0}\underbrace{\left[\underbrace{I_{LH}}_{<0}\left(\underbrace{I_{LL}}_{>0} + \underbrace{I_{HH}}_{>0}\right)\right]}_{<0}}_{<0} < 0\,.$$

Since Assumption (A2.8) is fulfilled for the base case set of parameters *a*≈*0.163, b*≈*0.024, g*≈*0.062, s=0.61, q=7.0,* τ*=50,000* and the Hopf critical values (listed in Table 2.1), respectively, there exists a surface of periodic solutions in the center manifold which has quadratic tangency with the eigenspace of $\lambda_{1,2}(\nu_c)$ for any of the system parameters ν. These periodic solutions are stable limit cycles where following rules generally hold (see Strogatz, 1994, p.251):

---

14 $\omega_c = \sqrt{g\left(a + b - s\exp\left[-q\frac{b}{g}\right]\right)}$

15 $A = \frac{1}{16}\left[\varphi_{LLL} + \varphi_{LHH} + \psi_{LLH} + \psi_{HHH}\right] + \frac{1}{16\omega_c}\left[\varphi_{LH}\left(\varphi_{LL} + \varphi_{HH}\right) - \psi_{LH}\left(\psi_{LL} + \psi_{HH}\right) - \varphi_{LL}\psi_{LL} + \varphi_{HH}\psi_{HH}\right]$

For the definition of the cubic coefficient for higher-dimensional systems see Guckenheimer and Holmes, 1983, p.152ff.

- The size of the stable limit cycle grows continuously from zero, and increases proportional to $\sqrt{\nu - \nu_c}$, for ν close to ν_c.
- The frequency of the limit cycle is given approximately by ω_c. This formula is exact at the birth of the limit cycle, and correct within $O(\nu - \nu_c)$ for ν close to ν_c. Therefore the period T of the cycle is given by

$$T = \frac{2\pi}{\omega_c} + O(\nu - \nu_c)\,. \tag{A2.9}$$

A.2.4 Calculation of Δ

The level of treatment, historically pursued in the US (and reflected by the flow rate *g*), is 31% of heavy users have received treatment each year and 13% of those treated leave heavy use. Hence, the outflow rate can be viewed as being

$$g = 0.062 + 0.13\,(u - 0.31) \tag{A2.10}$$

where *u* is the proportion of heavy users receiving treatment each year. Assume that we increase that proportion $u \rightarrow \tilde{u}$ and the outflow rate $g \rightarrow \tilde{g}$, respectively, according to Equation (A2.10). We define total discounted consumption, for a fixed outflow rate g, over a relevant finite planning horizon such as 50 years, starting with any combination of light and heavy users,

$$Q_{50_g} := \int_0^{50} e^{-0.04\,t}\left(16.42\,L_g(t) + 118.93\,H_g(t)\right) dt$$

(Following Everingham and Rydell, we assume light and heavy users consume at a fixed annual rate of 16.42 grams and 118.93 grams per year, respectively, and discount at 4% per annum) and we consider the change in discounted consumption in grams due to additional treatment, denoted by $Q_{50_g} - Q_{50_{\tilde{g}}}$.

Discounted cost of additional treatment ($u \rightarrow \tilde{u}$) over the same finite planing horizon is calculated as follows

$$Cost_{50_g} = C_{treat} \underbrace{\int_0^{50} e^{-0.04t} uH_g(t)\, dt}_{:= C_{50_g}}$$

where C_{treat} reflects the cost of a heavy user's treatment. To achieve the increment in the proportion of heavy users receiving treatment each year, $u \to \tilde{u}$, and the outflow rate $g \to \tilde{g}$, respectively, one has to provide an additional treatment budget, $C_{treat}\left(C_{50_{\tilde{g}}} - C_{50_g}\right)$. Therefore Δ may be interpreted as the averted consumption in grams per additionally treated heavy user

$$\mathrm{d} = \frac{\mathrm{k}}{C_{treat}} \underbrace{\left(\frac{Q_{50_g} - Q_{50_{\tilde{g}}}}{C_{50_{\tilde{g}}} - C_{50_g}} \right)}_{:= \Delta} \qquad \text{(A2.11)}$$

where κ measures the reduction in social cost per gram of consumption averted. Then a policy maker would have to estimate the cost of treating a user, C_{treat}, and judge the benefit of reducing cocaine consumption by a certain amount Δ. Condition (A2.11) determines how many grams of consumption an additional treatment must avert in order for additional treatment to be cost-effective ($\mathrm{d} \geq 1$).

Appendix to Chapter 3

A.3.1 Concavity of the Hamiltonian with Respect to Control u (Optimal Allocation Problem (3.8))

The first order derivative of the Hamiltonian $\mathcal{H}$ with respect to control u set equal to zero leads to the expression

$$\pi_2 = \underbrace{\frac{I\psi_u}{H\beta_u}}_{>0}\pi_1 . \tag{A3.1}$$

Therefore the signs of the costates are equal. Using (A3.1), the second order derivative of the Hamiltonian $\mathcal{H}$ with respect to u simplifies to:

$$\mathcal{H}_{uu} = \underbrace{\left(\underbrace{I\psi_{uu}}_{>0} - \underbrace{\frac{I\psi_u \beta_{uu}}{\beta_u}}_{<0} \right)}_{>0} \pi_1 \tag{A3.2}$$

Consequently the Hamiltonian $\mathcal{H}$ is concave with respect to the control u iff the costate π_1 is negative.

A.3.2 The Canonical System and the Local Stability Properties of its Equilibrium State (Model (3.8))

Due to the choice of the functional forms of initiation, prevention and treatment the optimal interior control $u^* = u^*(L, H, \pi_1, \pi_2)$ is implicitly defined by the Hamiltonian maximizing condition $\mathcal{H}_u = I(L,H)\psi_u(L,H,u^*)\pi_1 - H\beta_u(H,u^*)\pi_2 = 0$. We simplify $I(L,H)$, $\psi(L,H,u^*)$, and $\beta(H,u^*)$ to I, ψ, and β. Then the canonical system of the optimal control problem (3.8) in terms of states and costates is given by:

$$
\begin{aligned}
&\dot{L} = I\mathsf{y} - (a+b)L \\
&\dot{H} = bL - (g+\mathsf{b})H \qquad\qquad (A3.3)\\
&\dot{\mathsf{p}}_1 = 16.42(\mathsf{k}+\mathsf{g}) + (r+a+b-I_L\mathsf{y} - I\mathsf{y}_L)\mathsf{p}_1 - b\mathsf{p}_2 \\
&\dot{\mathsf{p}}_2 = 118.93(\mathsf{k}+\mathsf{g}) - (I_H\mathsf{y} + I\mathsf{y}_H)\mathsf{p}_1 + (r+g+\mathsf{b}+\mathsf{b}_H H)\mathsf{p}_2
\end{aligned}
$$

The limiting transversality conditions for the costates, given by (A3.4) and (A3.5), obviously hold since states and costates approach a stable state.

$$\lim_{t\to\infty} e^{-rt}\mathsf{p}_1(t)L(t) = 0 \qquad (A3.4)$$

$$\lim_{t\to\infty} e^{-rt}\mathsf{p}_2(t)H(t) = 0 \qquad (A3.5)$$

The interior stationary solution $\hat{E}_2$ follows from the simultaneous solution of the following set of equations:

$$
\left.\begin{aligned}
I\mathsf{y}_u\hat{\mathsf{p}}_1 - \hat{H}\mathsf{b}_u\hat{\mathsf{p}}_2 &= 0 \\
I\mathsf{y} - (a+b)\hat{L} &= 0
\end{aligned}\right\} \Rightarrow \hat{H}, \hat{u}^*
$$

$$\hat{L} = \frac{g+\mathsf{b}}{b}\hat{H} \quad > 0 \qquad (A3.6)$$

$$\hat{\mathsf{p}}_1 = -\frac{16.42(\mathsf{k}+\mathsf{g}) - b\hat{\mathsf{p}}_2}{r+a+b-I_L\mathsf{y} - I\mathsf{y}_L}$$

$$\hat{\mathsf{p}}_2 = -\frac{118.93(\mathsf{k}+\mathsf{g}) - (I_H\mathsf{y} + I\mathsf{y}_H)\hat{\mathsf{p}}_1}{r+g+\mathsf{b}+\mathsf{b}_H\hat{H}}$$

$$\hat{w}^* = \mathsf{g}\left(16.42\,\hat{L} + 118.93\,\hat{H}\right) - \hat{u}^* \quad > 0$$

A result of Dockner (1985) allows to characterize the local stability properties of the equilibrium of the canonical equations (A3.3). I.e. one can explicitly calculate the eigenvalues $\zeta_{j,\, j=1,2,3,4}$ of the Jacobian, denoted by J, of the nonlinear canonical equations:

$$\zeta_{1,2,3,4} = \frac{1}{2}\left(\rho \pm \sqrt{\rho^2 - 2K \pm 2\sqrt{K^2 - 4\|J\|}} \right)$$
$$K = K_1 + K_2 + K_3, \qquad \text{(A3.7)}$$

where:

$$K_1 = \begin{Vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial \pi_1} \\ \frac{\partial \dot{\pi}_1}{\partial L} & \frac{\partial \dot{\pi}_1}{\partial \pi_1} \end{Vmatrix}, \; K_2 = \begin{Vmatrix} \frac{\partial \dot{H}}{\partial H} & \frac{\partial \dot{H}}{\partial \pi_2} \\ \frac{\partial \dot{\pi}_2}{\partial H} & \frac{\partial \dot{\pi}_2}{\partial \pi_2} \end{Vmatrix}, \; K_3 = \begin{Vmatrix} \frac{\partial \dot{L}}{\partial H} & \frac{\partial \dot{L}}{\partial \pi_2} \\ \frac{\partial \dot{\pi}_1}{\partial H} & \frac{\partial \dot{\pi}_1}{\partial \pi_2} \end{Vmatrix}.$$

K_1 denotes the determinant of the Jacobian of the hypothetical one-state variable control for the state L and its corresponding adjoint variable π_1. The interpretation of K_2 is similar to K_1 if applied to the second state variable H and its costate π_2. The determinant K_3 measures the interactions. According to Feichtinger, Novak and Wirrl (1994) we are able to classify the local stability properties of equilibrium (A3.6) according to K and $\|J\|$:

Table A3.1: Classification of equilibria according to K and $\|J\|$

Inequalities	Equilibrium of the linearized canonical system
$K<0$ $0<\|J\| \le \left(K/2\right)^2$	Saddlepoint stability, real roots, two are negative and two are positive: local monotonicity
$\|J\| > \left(K/2\right)^2$ $\|J\| > \left(K/2\right)^2 + \rho^2\, K/2$	Saddlepoint stability, complex roots, two having negative real parts: transient oscillations
$\|J\| < 0$	One negative real eigenvalue, all other eigenvalues have positive real parts: instability except for the one-dimensional manifold
$\|J\| = \left(K/2\right)^2 + \rho^2\, K/2$ $K > 0, \|J\| > 0$	Two purely imaginary eigenvalues

$\|J\| > \left(K/2\right)^2$ $\|J\| < \left(K/2\right)^2 + r^2\, K/2$	Complex eigenvalues, positive real parts: locally unstable spirals

A.3.3 Concavity of the Hamiltonian with Respect to Controls *w* and *u* (Unconstrained Problem (3.11))

The Hamiltonian maximizing conditions lead to the expressions

$$\pi_1 = \frac{1}{I\psi_w} < 0 \quad \text{and} \quad \pi_2 = -\frac{1}{H\beta_u} < 0 . \tag{A3.8}$$

Therefore the signs of both costates are clearly defined. Using (A3.8), the second order derivative of the Hamiltonian $\mathcal{H}$ with respect to both controls is given by:

$$\mathcal{H}_{ww} = \underbrace{I}_{>0}\underbrace{\psi_{ww}}_{>0}\underbrace{\pi_1}_{<0} < 0 \quad \text{and} \quad \mathcal{H}_{uu} = -H\underbrace{\beta_{uu}}_{<0}\underbrace{\pi_2}_{<0} < 0 \tag{A3.9}$$

Consequently the Hamiltonian $\mathcal{H}$ is concave with respect to both controls.

A.3.4 The Derivation of the Canonical System for Optimal Interior Controls and Policies at the Border of the Admissible Region (Model (3.11))

Following equations summarize the necessary conditions for the optimal policies w^* and u^*, all in current values, where $\mathcal{H}=\mathcal{H}\left(L,H,\pi_1,\pi_2,w,u\right)$ denotes the Hamiltonian, $\mathcal{L}=\mathcal{L}\left(L,H,\pi_1,\pi_2,w,u,\theta,\sigma\right)$ the Lagrangean and π_1 and π_2 the costate variables:

$$\mathcal{H} = -\kappa\left(16.42L + 118.93H\right) - u - w + \pi_1\left(I\psi - (a+b)L\right) + \pi_2\left(bL - (g+\beta)H\right) \tag{A3.10a}$$

$$\mathcal{L} = \mathcal{H}\left(L,H,\pi_1,\pi_2,w,u\right) + \theta w + \sigma u \tag{A3.10b}$$

$$w^* = \arg\max_w \mathcal{L} \tag{A3.11a}$$

$$u^* = \arg\max_u \mathcal{L} \tag{A3.11b}$$

The limiting transversality conditions for the costates, given by (A3.12) and (A3.13), obviously hold since states and costates approach a stable state.

$$\lim_{t\to\infty} e^{-rt}\pi_1(t)L(t) = 0 \tag{A3.12}$$

$$\lim_{t\to\infty} e^{-rt}\pi_2(t)H(t) = 0 \tag{A3.13}$$

Particularly we have to distinguish between (A) optimal interior policies for the expenses on prevention and treatment and (B) optimal interventions at the border of the admissible region:

(A) Optimal interior policies

$$\left(w^* = -\frac{1}{m}\ln\left[-\frac{1}{mk\pi_1 I}\right] > 0;\quad u^* = (H+\delta)\sqrt[d-1]{-\frac{H+\delta}{cd\pi_2 H}} > 0;\quad \theta = \sigma = 0 \right):$$

Since we can express both controls in terms of states and costates, and $\psi(w^*) = h - (m\pi_1 I)^{-1}$, $\psi'(w^*) = (\pi_1 I)^{-1}$, $\beta(u^*) = -u^*(d\pi_2 H)^{-1}$, and $\beta_H(u^*) = u^*(\pi_2 H(H+\delta))^{-1}$, the canonical system for optimal interior controls is given by:

$$\begin{aligned}
\dot{L} &= h\,I - (a+b)\,L - \frac{1}{m\pi_1}\\
\dot{H} &= b\,L - g\,H + \frac{u^*}{d\pi_2}\\
\dot{\pi}_1 &= 16.42\kappa + \frac{I_L}{mI} + \left(r+a+b-h\,I_L\right)\pi_1 - b\,\pi_2\\
\dot{\pi}_2 &= 118.93\kappa + \frac{I_H}{m\,I} - \frac{u^*}{d\,H} + \frac{u^*}{H+\delta} - \left(h\,I_H\right)\pi_1 + \left(r+g\right)\pi_2
\end{aligned} \tag{A3.14}$$

(B) Policies at the border of the admissible region:

B.1 $\left(w^* = 0;\quad u^* = (H+\delta)\sqrt[d-1]{-\frac{H+\delta}{cd\pi_2 H}} > 0;\quad \theta = 1 + mk\pi_1 I;\ \sigma = 0 \right)$

Since we can express treatment, u^*, in terms of states and costates, and $\psi(w^*) = h - (m\pi_1 I)^{-1}$, $\psi'(w^*) = (\pi_1 I)^{-1}$, $\beta(u^*) = -u^*(d\pi_2 H)^{-1}$, and $\beta_H(u^*) = u^*(\pi_2 H(H+\delta))^{-1}$, the canonical system for optimal control at the border of the admissible region is given by:

$$
\begin{aligned}
\dot{L} &= I-(a+b)L \\
\dot{H} &= b\,L-g\,H+\frac{u^*}{d\mathrm{p}_2} \qquad\qquad \text{(A3.15)}\\
\dot{\mathrm{p}}_1 &= 16.42\,\mathrm{k} + \left(r+a+b-I_L\right)\mathrm{p}_1-b\,\mathrm{p}_2 \\
\dot{\mathrm{p}}_2 &= 118.93\mathrm{k} - \frac{u^*}{d\,H} + \frac{u^*}{H+\mathrm{d}} - I_H\mathrm{p}_1 + \left(r+g\right)\mathrm{p}_2
\end{aligned}
$$

B.2 $\left(w^* = -\frac{1}{m}\ln\left[-\frac{1}{mk\mathrm{p}_1 I}\right] > 0;\quad u^* = 0;\quad \mathrm{q} = 0;\ \mathrm{s} = 1\right)$

Since we can express prevention, w^*, in terms of states and costates, and $\mathrm{y}\left(w^*\right) = h-\left(m\mathrm{p}_1 I\right)^{-1}$, $\mathrm{y}'\left(w^*\right) = \left(\mathrm{p}_1 I\right)^{-1}$, $\mathrm{b}\left(u^*\right) = -u^*\left(d\mathrm{p}_2 H\right)^{-1}$, and $\mathrm{b}_H\left(u^*\right) = u^*\left(\mathrm{p}_2 H\left(H+\mathrm{d}\right)\right)^{-1}$, the canonical system for optimal control at the border of the admissible region is given by:

$$
\begin{aligned}
\dot{L} &= h\,I-(a+b)L-\frac{1}{m\mathrm{p}_1} \\
\dot{H} &= b\,L-g\,H \qquad\qquad \text{(A3.16)}\\
\dot{\mathrm{p}}_1 &= 16.42\mathrm{k} + \frac{I_L}{mI} + \left(r+a+b-h\,I_L\right)\mathrm{p}_1 - b\mathrm{p}_2 \\
\dot{\mathrm{p}}_2 &= 118.93\mathrm{k} + \frac{I_H}{mI} - \left(h\,I_H\right)\mathrm{p}_1 + \left(r+g\right)\mathrm{p}_2
\end{aligned}
$$

The stationary solution $\hat{E}_3$ follows from the simultaneous solution of $\dot{L}=0$, $\dot{H}=0$, $\dot{\mathrm{p}}_1=0$, and $\dot{\mathrm{p}}_2=0$ for the corresponding canonical systems.

References

Alexander, J.C. and Yorke, J.A. (1978). „Global Bifurcations of Periodic Orbits“. *American Journal of Mathematics 100*, pp.262-292.

Behrens, D.A., Caulkins, J.P., Tragler, G., Haunschmied, J.L., and Feichtinger, G. (1997). „A Dynamical Model of Drug Initiation: Implications for Treatment and Drug Control“. Working Paper #213, Vienna University of Technology.

Caulkins, J.P. and Reuter, P. (1997). „Setting Goals for Drug Policy: Harm Reduction or Use Reduction“. *Addiction*, Vol. 92, No. 9, pp.1143-1150.

Dockner, E. (1985). „Local Stability Analysis in Optimal Control Problems with Two State variables“. In G. Feichtinger, Ed. *Optimal Control Theory and Economic Analysis 2*, Amsterdam: North Holland.

Everingham, S.S. and Rydell, C.P. (1994). „Modeling the Demand for Cocaine“. *MR-332-ONDCP/A/DPRC*, Santa Monica, CA: RAND.

Everingham, S.S., Rydell, C.P., and Caulkins, J.P. (1995). „Cocaine Consumption in the US: Estimating Past Trends and Future Scenarios“. *Socio-Economic Planning Sciences*, Vol. 29, No. 4, pp.305-314.

Feichtinger, G. and Hartl, R. (1986). „*Optimale Kontrolle Ökonomischer Prozeße - Anwendungen des Maximumprinzips in den Wirschaftswissenschaften*“. Berlin: deGruyter.

Feichtinger G., Novak, A., and Wirrl, F. (1994). „Limit Cycles in Intertemporal Adjustment Models“. *Journal of Economic Dynamics and Control 18*, pp. 353-380.

Guckenheimer, J. and Holmes, P. (1983). „*Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*“. New York, NY: Springer.

Leonard, D. and, Long, N.V. (1992). „*Optimal Control Theory and Static Optimization in Economics*“. Cambridge: CUP.

Marsden, J.E. and McCracken, M. (1976). „*The Hopf Bifurcation and its Applications*“. New York-Heidelberg-Berlin: Springer.

Musto, D.F. (1987). „*The American Disease*“. New Haven, CT: Yale University Press.

National Institute on Drug Abuse (1991). „*Overview of the 1991 National Household Survey of Drug Abuse*“. press release, December.

Office of National Drug Control Policy (1996). „*The National Drug Control Strategy: 1996*“. Washington, DC: The White House.

Office of National Drug Control Policy (1994). „*The National Drug Control Strategy*“. Washington, DC: The White House.

Rice, D.P., Kelman, S., Miller, L.S., and Dunmeyer, S. (1990). „*The Economic Costs of Alcohol and Drug Abuse and Metal Illness*“: 1985, San Francisco, CA: Institute for Health and Aging, University of California.

Rydell, C. P. (1997). „*The Dynamics of Cocaine Control Policy Analysis Based on Exploratory Modeling*“. Working paper.

Rydell, C.P., Caulkins, J.P., and Everingham, S.S. (1996). „Enforcement or Treatment? Modeling the Relative Efficacy of Alternatives for Controlling Cocaine“. *Operations Research 44(5)*, pp. 1-9.

Rydell, C.P. and Everingham, S.S. (1994). „Controlling Cocaine. Supply Versus Demand Programs“. *MR-331-ONDCP/A/DPRC*, Santa Monica, CA: RAND.

Stares, P.B. (1996). „*Global Habit: The Drug Problem in a Borderless World*“. Washington DC: Brookings Institute.

Strogatz, S.H. (1994). „Nonlinear Dynamics and Chaos“. Addison-Wesley.

Tragler, G., Caulkins, J.P., and Feichtinger G. (1997). „The Impact of Enforcement and Treatment on Illicit Drug Consumption“. Working Paper #212, Vienna University of Technology.

List of Figures

List of Tables