Optimal Dynamic Allocation of Treatment and Enforcement in Illicit Drug Control[1]

Gernot Tragler[2]

Jonathan P. Caulkins[3]

Gustav Feichtinger[4]

December 1997

[1]This research was partly financed by the Austrian Science Foundation (FWF) under contract No. P11711-SOZ ("Dynamic Law Enforcement"), the National Consortium on Violence Research, and by the U.S. 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.

[2]Vienna University of Technology, Department for Operations Research and Systems Theory; Argentinierstr. 8/1192, A-1040 Vienna, Austria; email: tragler@e119ws1.tuwien.ac.at

[3]Carnegie Mellon University, H. John Heinz III School of Public Policy and Management; 5000 Forbes Ave., Pittsburgh, PA 15213; and RAND Drug Policy Research Center; email: caulkins+@andrew.cmu.edu

[4]Vienna University of Technology, Department for Operations Research and Systems Theory; Argentinierstr. 8/1192, A-1040 Vienna, Austria; email: gustav@e119ws1.tuwien.ac.at

Abstract

There has been considerable debate about what share of drug control resources should be allocated to treatment vs. enforcement. Most of the debate has presumed that there is one particular answer to that question, but it seems plausible that the optimal mix of interventions might vary as the size of the problem changes. We formulate the choice between drug treatment and enforcement as an optimal control problem. This perspective suggests that, if the initial number of drug users is small and tends to grow toward a larger equilibrium, then it is optimal to rely primarily on enforcement at first, in order to keep prices high and suppress initiation to the extent possible. Enforcement spending should increase as the number of users grows, but not nearly so fast in percentage terms as treatment spending should. Hence, it is optimal for treatment to receive a larger share of control resources when a drug problem is mature than when it is first growing. The model generates a variety of other insights, including: (1) detecting the onset of a drug epidemic quickly is valuable because total costs are much lower if control begins early, when the number of users is small; (2) people who perceive drug use to be costly for society should favor greater drug control spending per user and allocating a greater proportion of that spending to enforcement; (3) sharp price declines, such as those observed in the 1980s for cocaine in the U.S., do not necessarily imply a policy failure; indeed, it can be optimal to have such declines; and (4) under certain conditions it can be optimal to spend a very large amount per user – primarily on treatment – in a manner that prevents the epidemic from ever expanding beyond its initial, low use state.

Contents

Chapter 1

Introduction

Illicit drugs pose serious challenges for societies around the world. There are a wide range of drug control interventions, including prevention, treatment, and various forms of enforcement, among others. Not surprisingly there has been an energetic debate concerning how generously these different interventions should be funded. The most common manifestation of this debate concerns the effectiveness and relative roles of drug enforcement and treatment.

Different parties make different claims. Treatment is seven times more cost-effective than enforcement (Rydell & Everingham, 1994). Treatment is five times as effective as interdiction (Caulkins et al., 1997, updating Rydell & Everingham, 1994). Or, treatment and enforcement are both vital parts of an integrated drug control program (ONDCP, 1997).

Yet essentially all these claims are static. Few argue that, for example, enforcement should play the dominant role early in a drug epidemic, but relatively more resources should be allocated to treatment later on. And fewer still back such claims with rigorous analysis. But drug use and drug-related problems change substantially over time. There are waves of greater or lesser drug use, and the very use of a term like "epidemic" connotes a dynamic process. Hence, it seems plausible that there is no single, static answer to the question, "which is more effective, treatment or enforcement?" Rather, it may be more fruitful to ask, "how should the allocation of resources to treatment and enforcement vary over the course of a drug epidemic?" The goal of this paper is to ask and at least partially answer the latter question by applying the tools of optimal control theory to a simple dynamic model of drug use.

This paper is not the definitive treatise on the topic of dynamic drug control. The underlying dynamic model is very simple, using just one state variable to reflect drug use. It does not differentiate between different intensities of use or different drugs. The controls are restricted to treatment and enforcement, and enforcement's effects are modulated through prices. I.e., the model captures the "risks and prices" (Reuter & Kleiman, 1986) vision of drug enforcement, but not its ability to raise "search times" (Moore, 1973; Kleiman, 1988) or reduce the harms per unit use by controlling the negative externalities associated with drug markets and drug use. Hence, this paper should be viewed as a first step into an important topic area, and it should be considered in conjunction with parallel efforts (e.g., Behrens et al., 1997a,b) and subsequent research. Nevertheless, the model produces important insights for how drug policy should be pursued and evolve over time.

The model parameter values are inspired by evidence from the U.S. cocaine epidemic of the 1970s – 1990s, but should not be construed as precise estimates for three reasons. First, data with which these parameters can be estimated are limited. Second, the model is too simple to capture the rich details of an actual drug epidemic. Third, our goal is not to describe the U.S. cocaine epidemic but rather to derive qualitative insights that might generalize to other drugs in other contexts. Occasionally we make reference to specific numerical findings where appropriate, but those should be understood as merely being suggestive.

Chapter 2

The Model

2.1 The Base Model

We consider a continuous time optimization problem where the decision maker (a reified representation of "the government") seeks to minimize the discounted stream of the sum of social costs caused by the use of an illicit drug and the monetary costs resulting from controls used to fight the drug problem. The controls to be considered in our model are price-raising enforcement (i.e., enforcement against dealers), $v(t)$, and treatment of drug users, $u(t)$. Strictly speaking, $v(t)$ and $u(t)$ denote the budgets for enforcement and treatment spending at time t, respectively. Throughout this paper we will assume that the non-negativity condition

$$u(t) \geq 0, v(t) \geq 0 \; \forall t \tag{2.1}$$

is satisfied.

There are many effects of enforcement against dealers. For one, it leads to the apprehension of dealers, of course. This effect, however, will not be considered in this paper, which implicitly assumes that incarcerated dealers can be replaced and users can find alternative sources relatively easily. This allows us to concentrate on only one group of people, i.e. the number of users, $A(t)$ (state variable). Instead we focus on enforcement's ability to act like a tax that drives up the cost of distributing drugs (Reuter & Kleiman, 1986). As pointed out by Rydell & Everingham (1994), "[...] *the money spent on supply control causes increases in the cost to producers of supplying the* *cocaine. That increased cost of supply gets passed along to the consumer as price increases, which in turn causes the number of users to decline as inflows to cocaine use decrease and outflows increase."* This, of course, should be true for any (illicit) drug.

To express this idea in mathematical terms, we let the drug price, p, depend on $v(t)$. In addition, because of "enforcement swamping" (Kleiman, 1993) the impact of a given level of enforcement on a drug market depends on the size of that market so we let p also depend on $A(t)$ which is used as a measure of the drug market's size. We thus have

$$p = p\left(A(t), v(t)\right) = p\left(\frac{v(t)}{A(t)+\epsilon}\right), \tag{2.2}$$

where ϵ is a constant used to avoid zero denominators for $A(t) \to 0$ ($\epsilon > 0$). This constant has essentially no effect on the results and is included as a mathematical convenience. If there is no enforcement, the price is assumed to be a positive constant, d, i.e. we have

$$p\left(A(t), 0\right) = d. \tag{2.3}$$

When we say "no enforcement" we do not envision drug legalization, but rather whatever minimum level of enforcement is necessary for what Reuter calls the "structural consequences of product illegality" to obtain. Given the "enforcement swamping" image, it is reasonable to assume that

$$p_A := \frac{\partial p}{\partial A} < 0$$

and

$$p_v := \frac{\partial p}{\partial v} > 0.$$

Analogously, the proportion of users who cease use as a result of treatment, β, will not only depend on the money spent but also on the number of those people who are potentially treated, i.e.

$$\beta = \beta\left(A(t), u(t)\right) = \beta\left(\frac{u(t)}{A(t)+\delta}\right) \tag{2.4}$$

with a positive constant δ included to avoid zero denominators (cf. ϵ in the price function above). We include a "natural", untreated exit term explicitly below, so we assume that

$$\beta\left(A(t), 0\right) = 0. \tag{2.5}$$

We also assume that a given level of funding will lead a smaller proportion of users to quit if those funds are spread over more users, i.e.

$$\beta_A := \frac{\partial \beta}{\partial A} < 0,$$

and finally that more people can be treated if more resources are deployed:

$$\beta_u := \frac{\partial \beta}{\partial u} > 0.$$

With (2.2) and (2.4) we are now able to specify the system dynamics (i.e., the rate of change of the number of users) as follows:

$$\dot{A}(t) = kp\left(A(t), v(t)\right)^a - c\beta\left(A(t), u(t)\right)^z A(t) - \mu p\left(A(t), v(t)\right)^b A(t), \quad (2.6)$$

where k, c, μ are proportionality constants (all > 0) representing initiation, treatment effectiveness, and natural desistance, respectively. $a < 0$ and $b > 0$ govern how price suppresses initiation and encourages desistance, respectively. If β were constant, $(a - b)$ would be the elasticity of the equilibrium number of users with respect to changes in price. So these two parameters will be referred to collectively as governing the elasticity of participation, which is the difference between the long run and short run elasticities of demand. The parameter $0 < z < 1$ reflects diminishing marginal efficiency of treatment. Such diminishing returns have been included in past models (e.g., Rydell & Everingham, 1994) and reflect "cream skimming." The image is that some users are easier to treat than others are, and that the treatment system has some capacity and incentive to focus efforts on individuals who are more likely to benefit from an intervention. As treatment funding and the number of people treated grows, the system can afford to be less and less selective.

The decision maker is assumed to care about both direct government expenditures $(u(t)+v(t))$ and social costs of drug consumption, so the utility functional, J, is given by

$$J = -\int_0^\infty e^{-rt}\left(\kappa A(t) p\left(A(t), v(t)\right)^\omega + u(t) + v(t)\right) dt. \quad (2.7)$$

The constant κ represents the social cost per gram of consumption, $\omega < 0$ denotes the short run elasticity of demand for current users, and $r > 0$ is the discount (time preference) rate. r is assumed to be constant over the whole planning horizon. If r is small (close to zero), then the government is interested in what happens in the future ("farsighted"), whereas a "myopic" government is characterized by rather big values of r. Note that we consider an infinite planning horizon; the case of a fixed finite planning horizon can easily be dealt with (see, e.g., Chapter 5).

The goal of the government is to minimize the present value of the total social costs which is defined by

$$\max_{u,v} J. \tag{2.8}$$

In this paper we will consider four variations of the base model defined above by (2.8), (2.7), and (2.6) which differ in the constraints placed on the controls. They are explained in the rest of this chapter. Where there is no danger of confusion, time arguments t will be omitted in what follows.

2.2 The Model without Controls

If the government is not aware of a drug problem, or simply does not want to spend money controlling it, we have

$$u(t) = v(t) = 0 \ \forall t$$

which means that we have no optimization problem and are left with a descriptive model. With (2.3) the utility functional (2.7) reduces to

$$J = -\int_0^\infty e^{-rt} \kappa A(t) d^\omega dt, \tag{2.9}$$

and – using (2.3) and (2.5) – the system dynamics (2.6) becomes

$$\dot{A} = kd^a - \mu d^b A. \tag{2.10}$$

In the model without controls we thus leave the users to their fate and just watch the system evolve, or – to put it in other words – we can find out what happens with the number of users and the total consumption if the government does not intervene (beyond the minimal level of enforcement necessary to make the drug market behave like an illegal market). This problem is not offered as either a descriptive or a prescriptive model, but rather as a counterfactual or foil against which the other models can be compared.

2.3 The Allocation Problem

Ideally the government would be free to choose whatever controls (u, v) minimize the objective function. Practically speaking, the political process and the limitations of human institutions are such that optimal controls can not always be pursued. One simplistic image is that society allocates resources to drug control not optimally but rather in proportion to how large the drug problem is. When the drug problem is perceived to be small, the public might resist having many tax dollars being spent on it; 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 be made. There are many possible measures of the severity of the drug problem, but to be consistent with the measure used in (2.2), we will use the number of users A, and consider the modifying the control problem above by adding the constraint that the drug control budget $B(t)$ must be proportional to the number of users, $A(t)$, i.e.,

$$u(t) + v(t) = B(t) = GA(t), \tag{2.11}$$

where G denotes the amount in monetary units the government is willing to spend per user at any time. We consider two different types of allocation problems, as described in the following two subsections.

2.3.1 Constant Fraction Control

The first method of choosing the levels of enforcement and treatment is easy to implement ("dumb") and thus has the advantage that implementation costs – which are neglected in all models presented here – are comparatively low. In particular, the government chooses that fraction of the total budget which is intended for enforcement, $0 < f < 1$, once and for all, i.e.

$$v(t) = fGA(t) \; \forall t$$

and

$$u(t) = GA(t) - v(t) = (1 - f)GA(t) \; \forall t.$$

The government may choose f optimally, i.e.

$$f = \arg\max_f J$$

with

$$J = -\int_0^\infty e^{-rt} \left(\kappa A(t) p\left(A(t), fGA(t)\right)^\omega + GA(t)\right) dt \tag{2.12}$$

subject to

$$\dot{A} = kp\left(A, fGA\right)^a - c\beta\left(A, (1-f)GA\right)^z A - \mu p\left(A, fGA\right)^b A. \tag{2.13}$$

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

2.3.2 Optimal Control

The other allocation problem to be considered in this paper is to solve (2.8) with (2.7) subject to the state dynamics (2.6) and the budget constraint (2.11) as a "standard" optimal control problem.

2.4 The Unrestricted ("Free") Control Problem

The fourth problem is the base model (2.8), (2.7), and (2.6) without the budget constraint (2.11), i.e. the controls enforcement and treatment are unrestricted except for the standard non-negativity assumption (2.1) that has to be met in all the models.

This last model might be more appropriate than one with a budget constraint such as (2.11) if treatment and enforcement resources are not allocated from a single pot. Drug enforcement might compete more with other forms of law enforcement for funding than it does with treatment. Likewise treatment may compete most directly with other forms of health and welfare spending. Also, even if one believed that drug control spending really is constrained, this problem provides an interesting comparison; it indicates how much could be gained by being able to invest heavily in drug control resources before a drug problem reaches a severe stage.

Chapter 3

Analysis

This chapter is devoted to the analysis of the four models presented above. In what follows, indices $_n$, $_c$, $_a$, and $_f$ associated with the state variable A, the control variables u and v, etc., indicate that these values refer to the uncontrolled (**no** control), **c**onstant fraction control, optimally controlled **a**llocation, and unrestricted (**f**ree) control problems, respectively. The numerical analysis has been performed with Stephen Wolfram's *Mathematica*-System (Wolfram, 1996). A detailed interpretation of the results as well as a comparison of the four cases will be the contents of the next chapter.

3.1 Specification of the Price and Treatment Functions

Since our aim is to obtain qualitative insight into the structure of optimal enforcement and treatment policies, respectively, we specify the price and treatment functions (2.2) and (2.4) as follows:

$$p(A, v) = d + e\frac{v}{A + \epsilon}, \tag{3.1}$$

where e measures the efficiency of enforcement[1], and

$$\beta(A, u) = \frac{u}{A + \delta}, \tag{3.2}$$

[1]Modeling price as linear in the ratio of enforcement to market size is consistent with the "enforcement swamping" model (Kleiman, 1993) and follows the lead of Caulkins et al. (1997).

i.e., linear functions in v and u, respectively.

3.2 Parameter Choices

The base values of the system parameters which will be used in the numerical analysis below are given in Table 3.1. Their derivation can be found in Appendix A.1. Note that the parameters d, e, k, κ, and μ appear twice in this table. The values in parantheses [ ], which will not be used for the analysis, are the "original" values as described in Appendix A.1. However, for technical convenience we reduce the number of parameters by setting $\kappa = 1$ and recalculate the values of d, e, k, and μ; this procedure is described in Appendix A.2.

In order to make understanding of the following results easier, we introduce a parameter κ_{pgc} to indicate that the associated analysis has been carried out with the value of the *p*er *g*ram *c*osts being \$ *pgc*. E.g., the base value of \$ 100 in social cost per gram of consumption is denoted by $\kappa_{pgc} = \kappa_{100}$. κ_{pgc} does not equal *pgc* because it subsumes the leading coefficient of the demand curve, i.e. the proportionality constant that makes consumption per user proportional to the price raised to the short term elasticity.

3.3 The Model without Controls

The model with $u(t) = v(t) = 0 \ \forall t$ – (2.9) and (2.10) – is a special case of the constant fraction control allocation problem – (2.12) and (2.13) – with $G = 0$. Therefore we immediately move on to the next section where the constant fraction control model is analysed.

3.4 The Allocation Problem

3.4.1 Constant Fraction Control

The state dynamics (2.13) is a first order linear differential equation which can easily be solved analytically (please note that we assume that $\delta = \epsilon = 0$ here; as we will see later, neither δ nor ϵ have any appreciable impact on the

Table 3.1: Base parameter values; the values in parentheses [ ] are the "original" values (calculated before the reduction of parameters) to be used in interpreting the results but they were not used for the analysis

Parameter	Base Value	Definition
a	-0.25	elasticity of initiation with respect to price
b	0.25	elasticity of quitting with respect to price
c	0.04323	efficiency of treatment
d	0.03175	price with minimal enforcement
$[d]$	[0.06792]	[price with minimal enforcement (in thousands of dollars per gram)]
δ	0.001	constant to avoid division by zero
e	0.01241	efficiency of enforcement
$[e]$	[0.02655]	[efficiency of enforcement]
ϵ	0.001	constant to avoid division by zero
G	1.6	drug control budget (in thousands of dollars) per addict
k	$472,618$	initiation proportionality constant (new users / year)
$[k]$	$[571,573]$	[initiation proportionality constant (new users / year)]
κ	1	per gram costs proportionality constant
$[\kappa]$	[1.46259]	[per gram costs proportionality constant]
μ	0.22786	outflow rate from use
$[\mu]$	[0.18841]	[outflow rate from use]
ω	-0.5	short run elasticity of demand
r	0.04	annual discount rate
z	0.6	$1 - z$ reflects extent of diminishing returns to treatment

results):

$$A_c(t) = \frac{k\Phi^a}{\Omega} + \left(A_0 - \frac{k\Phi^a}{\Omega}\right) e^{-\Omega(t-t_0)}\,,\ A_c\left(t_0\right) = A_0 \tag{3.3}$$

where

$$\Phi = d + efG,$$

$$\Psi = (1-f)\,G,$$

and

$$\Omega = c\Psi^z + \mu\Phi^b.$$

Looking at (3.3) and using the fact that $\Omega \geq 0$, we see that

$$\hat{A}_c := \lim_{t\to\infty} A_c(t) = \frac{k\Phi^a}{\Omega} = \hat{A}_c\left(f\right) \tag{3.4}$$

is the steady state to be approached, and by solving the integral of the utility functional (2.12) with (3.3) we get

$$J_c := -\frac{\kappa\Phi^\omega + G}{\Omega + r}\left(\frac{k\Phi^a}{r} + A_0\right) = J_c\left(A_0, f\right) \tag{3.5}$$

as the total social loss caused by the drug problem. Let further J_c^* and f_c^* be defined by

$$J_c^* := \max_f J_c\left(A_0, f\right) = J_c^*\left(A_0\right)$$

and

$$f_c^* := \arg\max_f J_c\left(A_0, f\right) = f_c^*\left(A_0\right),$$

respectively, i.e., f_c^* is that (optimal) fraction of the total budget that goes for enforcement.

Table 3.2 summarizes the values of $\hat{A}_c$, $\hat{u}_c$, $\hat{v}_c$, f_c^*, and J_c^*, respectively, for both the case of no control (in which case we have $G = 0$) and the base value of G, 1.6. It makes sense to impose controls on the drug problem even if those controls are "dumb", as the overall utility functional for the controlled case is greater (less negative) than that of the uncontrolled one. In other words, the damage caused by drug consumption is so high that it pays to spend a relatively large amount of money controlling the drug problem. From Fig. 3.1 (middle curve), the worst value of utility functional (3.5) with controls

Table 3.2: Steady state values $\hat{A}_c$, $\hat{u}_c$, $\hat{v}_c$, optimal fractions f_c^*, and optimal utility functional values J_c^*, both for $G = 0$ (no control) and $G = 1.6$ (base value) with $A_0 = 100,000$

G	$\hat{A}_c$	$\hat{u}_c$	$\hat{v}_c$	f_c^*	J_c^*
0	11,640,259	0	0	–	– 1,157,597,439
1.6	7,718,898	3,421,016	8,929,221	0.723	– 930,532,431

Figure 3.1: Utility functional values $J_c\,(A_0 = 100,000\,,\ f)$ as functions of f for per gram social costs \$ 50 (thin), \$ 100 (regular), and \$ 200 (thick), respectively; the vertical lines at f = 0.670, 0.723, and 0.754 indicate the optimal fractions f_c^* for the corresponding per gram costs \$ 50, \$ 100, and \$ 200, respectively

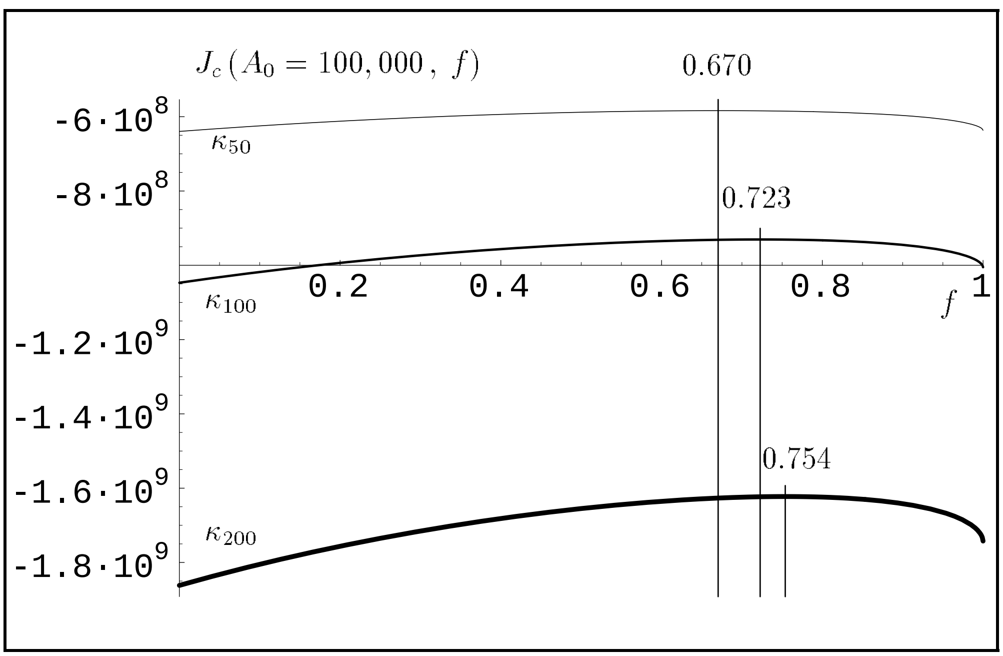


Table 3.3: Percent changes $e_{\hat{A}_n}$, $e_{\hat{J}_n}$, and $e_{J_n^*}$ of $\hat{A}_n$, $\hat{J}_n$, and J_n^* with respect to 1 % increases of the base parameter values ("elasticities")

Parameter	1.01· · (Base Value)	$e_{\hat{A}_n}$	$e_{\hat{J}_n}$	$e_{J_n^*}$
a	-0.2525	0.866	0.866	0.863
b	0.2525	0.866	0.866	0.610
c	0.04366	0.000	0.000	0.000
d	0.03207	-0.496	-0.990	-0.917
e	0.01254	0.000	0.000	0.000
k	$477,344$	1.000	1.000	0.996
κ	1.01	0.000	1.000	1.000
μ	0.23014	-0.990	-0.990	-0.701
ω	-0.505	0.000	1.740	1.740
r	0.0404	0.000	0.000	-1.277
z	0.606	0.000	0.000	0.000

(namely $-1,047,004,408$, which occurs for $f = 0$, or trying to control drug use only with treatment) is still better than what we get with no drug control spending.

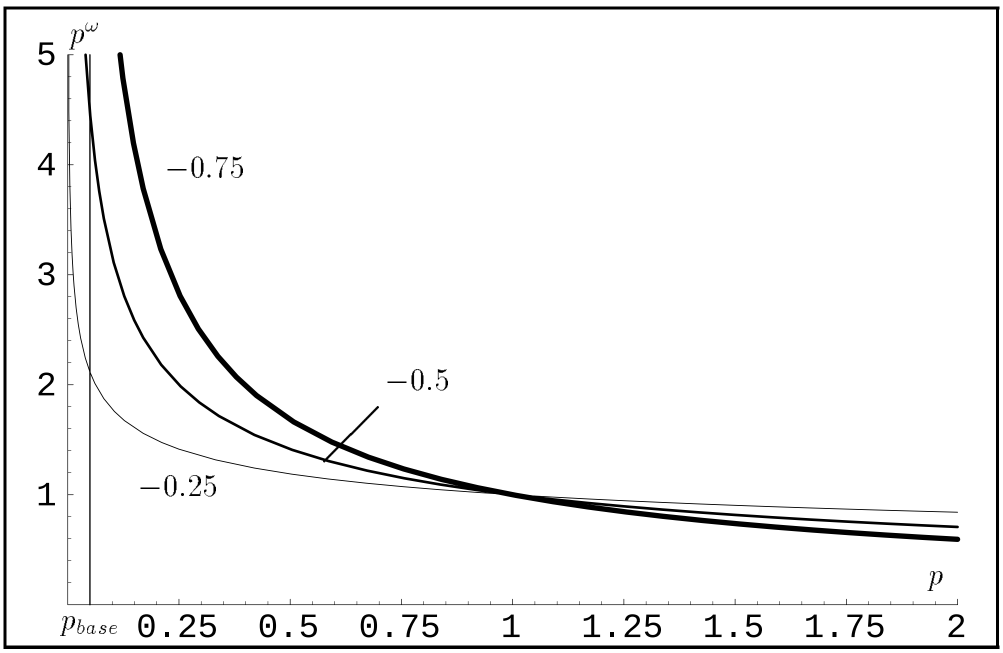
In Tables 3.3 (no control) and 3.4 (constant fraction control) we find the effects of 1 % changes of the base parameter values on the steady state values $\hat{A}$ and the utility functional values $\hat{J}$ (at steady state) and J^* (overall). Additionally, in Table 3.4 we have columns for the elasticities of the controls at steady state, $\hat{u}_c$ and $\hat{v}_c$, as well as enforcement's share at steady state, $\hat{f}_c$, which don't exist in the case of no control.

Before interpretating these values, please note that – due to the scale which has been chosen for the computation of the parameter values – the influence of changes in the elasticities is opposite to what one would expect (follow Fig. 3.2). Assume first the price p to be at its base value, $p_{base} = 0.04989$. This is clearly less than 1, and it would take more than a 5,000 % increase in per capita enforcement spending, $\frac{v}{A(+\epsilon)}$, for it to exceed 1. For values of p which are less than 1, however, an increase in the absolute value of the short term elasticity of demand ω, e.g., does not lead to a decrease but

Table 3.4: Percent changes $e_{\hat{A}_c}$, $e_{\hat{u}_c}$, $e_{\hat{v}_c}$, $e_{\hat{f}_c}$, $e_{\hat{J}_c}$, and $e_{J_c^*}$ of $\hat{A}_c$, $\hat{u}_c$, $\hat{v}_c$, $\hat{f}_c$, $\hat{J}_c$, and J_c^* with respect to 1 % increases of the base parameter values ("elasticities")

Parameter	1.01· · (Base Value)	$e_{\hat{A}_c}$	$e_{\hat{u}_c}$	$e_{\hat{v}_c}$	$e_{\hat{f}_c}$	$e_{\hat{J}_c}$	$e_{J_c^*}$
a	-0.2525	0.809	0.251	1.024	0.212	0.785	0.769
b	0.2525	0.577	1.155	0.356	-0.220	0.603	0.472
c	0.04366	-0.302	1.201	-0.879	-0.579	-0.235	-0.155
d	0.03207	-0.378	0.674	-0.783	-0.406	-0.586	-0.530
e	0.01254	-0.058	-1.288	0.416	0.472	-0.228	-0.241
G	1.616	-0.251	0.620	0.795	0.048	-0.117	-0.077
k	$477,344$	1.000	0.998	1.001	0.001	1.000	0.996
κ	1.01	0.014	-0.196	0.095	0.081	0.749	0.744
μ	0.23014	-0.715	-1.887	-0.265	0.453	-0.767	-0.610
ω	-0.505	0.077	-1.070	0.517	0.440	1.180	1.154
r	0.0404	0.021	-0.297	0.143	0.122	0.007	-1.216
z	0.606	0.031	1.023	-0.349	-0.380	0.075	0.075

Figure 3.2: p^ω as a function of p for $\omega = -0.25$ (thin), -0.5 (regular), and -0.75 (thick); the vertical line at $p = 0.04989$ indicates the base value of p, p_{base}



rather to an increase in consumption. The same is true for the elasticities of participation, a and b, and the argument also applies to the parameter z at the treatment term of the state equation, where we have $\beta_{base} = 0.13846 < 1$.

We skip over the interpretation of the elasticities for the case of no control (the values can be derived directly from (3.4) and (3.5) by setting $G = 0$, and the interpretation is analogous to that of the constant fraction control problem).

If we impose the constant fraction control (follow Table 3.4), all values – except for $\hat{f}_c$, which makes sense – increase more or less proportionally with k. Secondly, taking the sum of the absolute values of the elasticities as a measure, the strongest effects – apart from k – come with changes of μ, whereas increasing r by 1 % has nearly no impact. Finally, the variable that is influenced strongest is treatment at steady state, $\hat{u}_c$.

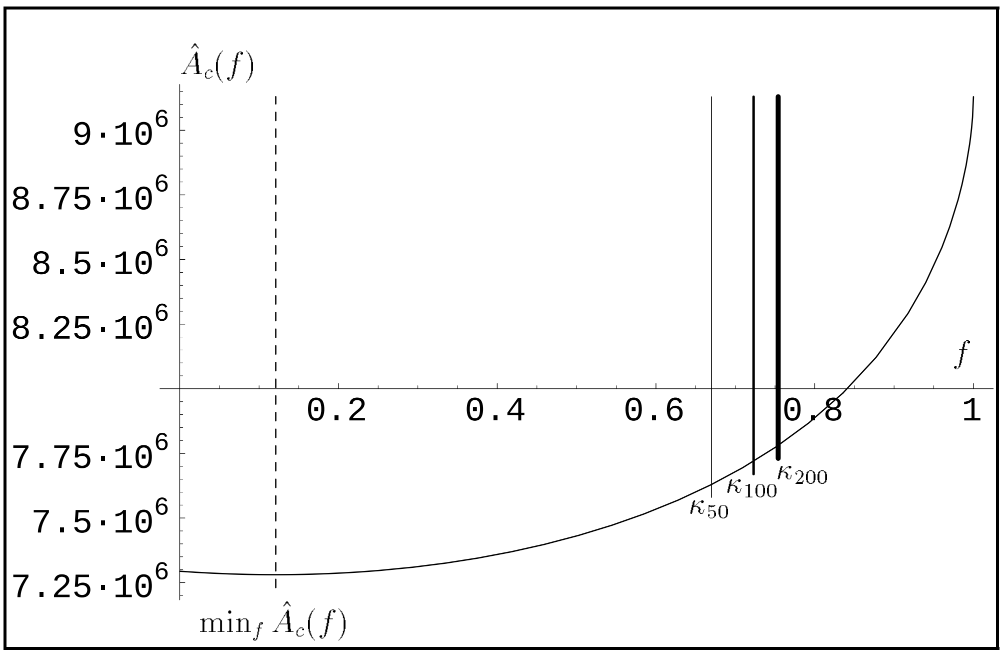
Sensitivity analysis with respect to the social (per gram) costs reveals most interesting insights (follow Figs. 3.1 and 3.3). The higher the costs, the greater the share the government optimally allocates to enforcement (Fig. 3.1), which brings up prices and reduces consumption per user even though shifting resources out of treatment increases the number of users slightly (Fig. 3.3). This is a parallel to the difference between the U.S.A. and the Netherlands: the U.S.A. have high perceived cost per unit use of heroin, high heroin prices, low consumption per user and slightly more users per capita than the Netherlands.

Before we move on to the optimally controlled allocation problem, please note that

$$\begin{aligned}
&\hat{A}_c\,(f_c^*) = \\
&= \hat{A}_c \begin{pmatrix} 0.670 \\ 0.723 \\ 0.754 \end{pmatrix} = \left\{ \begin{array}{ll} 7,629,655 & (\kappa_{50}) \\ 7,718,898 & (\kappa_{100}) \\ 7,780,723 & (\kappa_{200}) \end{array} \right\} > 7,280,669 = \hat{A}_c\,(0.121) = \\
&\qquad = \min_f \hat{A}_c(f),
\end{aligned}$$

which means that the steady state values at the optimal fractions f_c^* are greater than the minimum steady state value $\min_f \hat{A}_c(f)$ (see Fig. 3.3). In other words, minimizing the number of users does not minimize the total cost to society.

Figure 3.3: Steady state value $\hat{A}_c(f)$ as a function of enforcement's share f; the dashed vertical line at $f = 0.121$ indicates that fraction f which minimizes the steady state value $\hat{A}_c(f)$; the vertical lines at $f = 0.670$ (thin), 0.723 (regular), and 0.754 (thick) indicate the optimal fractions f_c^* and the corresponding steady state values $\hat{A}_c(f_c^*)$ for per gram costs \$ 50, \$ 100, and \$ 200, respectively



3.4.2 Optimal Control

The optimal control allocation problem as well as the unrestricted control problem are solved with the help of the maximum principle (see, e.g., Feichtinger & Hartl, 1986, or Leonard & Long, 1992). For simplicity, in what follows we abbreviate the functions $\beta(A,u)$ and $p(A,v)$ by β and p, respectively.

The current value Hamiltonian H for the problem (2.8) with (2.7) subject to (2.6) and (2.11) is given by

$$H = -\kappa A p^{\omega} - GA + \lambda \left(k p^{a} - c\beta^{z} A - \mu p^{b} A \right) \tag{3.6}$$

where λ denotes the current value costate variable. Note that according to (2.11) we can replace u by

$$u = GA - v$$

so that (2.8) is reduced to a maximization with respect to only one variable, i.e., v.

Applying Pontryagin's maximum principle to (3.6), we derive the necessary optimality condition

$$v = \arg\max_{v} H. \tag{3.7}$$

A sufficient condition for H to be concave with respect to the control v is the negativity of the costate variable, i.e.

$$\lambda < 0 \Rightarrow H_{vv} < 0$$

(for a proof see Appendix A.3). Although we cannot prove analytically that $\lambda < 0$, the numerical analysis shows that this condition is met, so we may solve (3.7) by claiming $H_v = 0$, which leads to

$$\lambda = \frac{\kappa\omega p^{\omega-1} p_v A}{a k p^{a-1} p_v - c z \beta^{z-1} \beta_v A - \mu b p^{b-1} p_v A} =: \frac{Z(\lambda)}{N(\lambda)}. \tag{3.8}$$

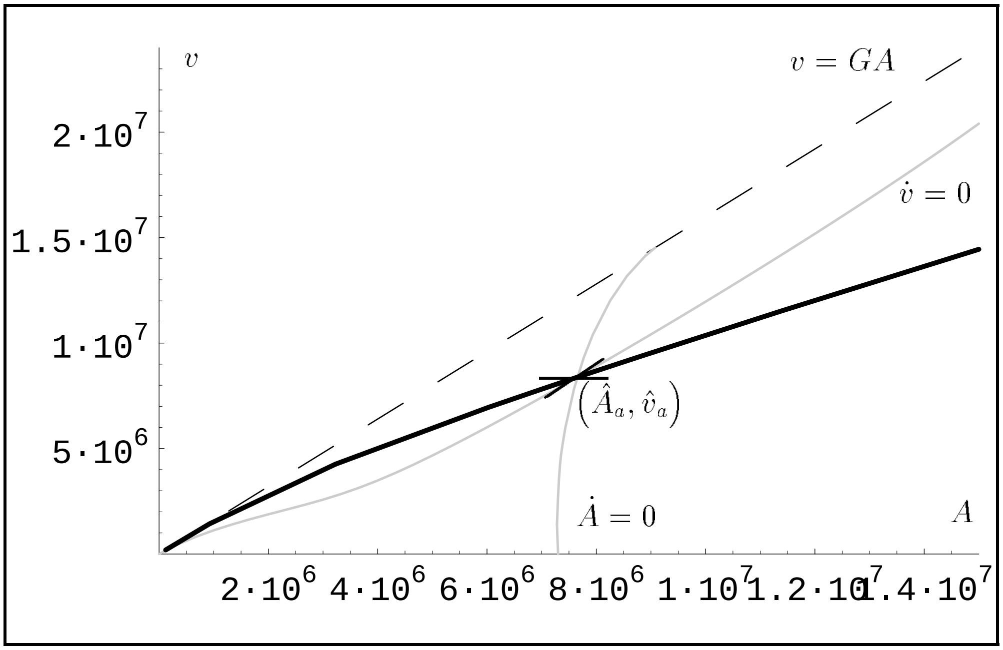
The costate equation is given by

$$\dot{\lambda} = r\lambda - H_A \tag{3.9}$$

which is used together with (3.8) to derive a differential equation for the control v which is given by (for details, see Appendix A.4)

$$\dot{v} = \frac{N(\lambda)\left[Z(\lambda)\left(r - H_A^{(2)} \right) - N(\lambda) H_A^{(1)} \right] - Z(\lambda_A)\dot{A}}{Z(\lambda_v)} \tag{3.10}$$

Figure 3.4: Phase portrait in the A-v-plane for the optimally controlled allocation problem using the base parameter values from Table 3.1; the dashed line is the budget constraint border line $v = GA$, the grey curves represent the $\dot{A} = 0$ and $\dot{v} = 0$ isoclines, respectively, and the black curves are the stable manifolds of the saddle point equilibrium $\left(\hat{A}_a, \hat{v}_a\right)$ (intersection of the isoclines)



with $N(\lambda)$ and $Z(\lambda)$ from (3.8) and the other expressions as given in Appendix A.4.

Note that the usual (Mangasarian) sufficiency conditions are not satisfied, since the maximized Hamiltonian is not concave with respect to the state variable.

The steady state values for this problem are given by simultaneously solving $\dot{A} = 0$ and $\dot{v} = 0$. Using the base parameter values from Table 3.1, these two curves in the A-v-plane are given in Fig. 3.4 (grey curves). Their intersection is a saddle point equilibrium $\left(\hat{A}_a, \hat{v}_a\right)$ (see Table 3.5), and the optimal trajectories are given by the two stable manifolds (black curves). This implies that for initial states above the steady state value ($A(0) >$

Table 3.5: Steady state values $\hat{A}_a$, $\hat{u}_a$, and $\hat{v}_a$, enforcement's share at steady state, $\hat{f}_a$, and optimal utility functional value, J_a^*, in the optimal control allocation problem

$\hat{A}_a$	$\hat{u}_a$	$\hat{v}_a$	$\hat{f}_a$	J_a^*
$7,652,824$	$3,857,023$	$8,387,495$	0.685	$-912,889,670$

$\hat{A}_a$), the optimal treatment and enforcement rates are greater than their equilibrium levels, but gradually decrease to these levels, driving the number of users to its long-run steady state. Analogously, for initial states below the steady state value ($A(0) < \hat{A}_a$), the optimal treatment and enforcement rates are relatively low and gradually increase to $\hat{u}_a$ and $\hat{v}_a$, respectively, while $A(t)$ converges to $\hat{A}_a$.

Table 3.6 is a summary of the effects of 1 % changes of the base parameter values on the steady state values $\hat{A}_a$, $\hat{u}_a$, and $\hat{v}_a$, as well as enforcement's share at steady state, $\hat{f}_a$, and the utility functional values $\hat{J}_a$ (at steady state) and J_a^* (overall); cf. Tables 3.3 and, especially, 3.4.

Neither changes in δ nor in ϵ affect the equilibrium values, so introducing these mathematical artifices does not jeopardize the integrity of the model. Secondly, there is a striking similarity of the elasticity values of this problem and those of the other – constant fraction – allocation problem (cf. Tables 3.4 and 3.6). Again, all values – except for $\hat{f}_a$ – increase proportionally with k ($e_{J_a^*}$ deviates negligibly); again, the variable that is influenced most strongly is treatment at steady state, $\hat{u}_c$; and again, increasing r by 1 % has nearly no impact. However, here it is ω that causes the strongest effects (apart from k), although μ is fairly close.

Sensitivity analysis with respect to the social cost per gram (κ) is particularly important because these costs are not only difficult to measure in an objective sense, they are also inherently subjective because different people may wish to include or exclude different costs in a social planner's objective function. For instance, some but not all people would want to count intangible "pain and suffering" costs associated with drug-related crime. To find out how different estimations of the per gram costs affect the outcomes, in Figs. 3.5 – 3.8 we have the steady state values $\hat{A}_a$ (Fig. 3.5), enforcement's shares at steady state $\hat{f}_a$ (Fig. 3.6), and the optimal utility functional values

Table 3.6: Percent changes $e_{\hat{A}_a}$, $e_{\hat{u}_a}$, $e_{\hat{v}_a}$, $e_{\hat{f}_a}$, $e_{\hat{J}_a}$, and $e_{J_a^*}$ of $\hat{A}_a$, $\hat{u}_a$, $\hat{v}_a$, $\hat{f}_a$, $\hat{J}_a$, and J_a^* with respect to 1 % increases of the base parameter values ("elasticities")

Parameter	1.01· · (Base Value)	$e_{\hat{A}_a}$	$e_{\hat{u}_a}$	$e_{\hat{v}_a}$	$e_{\hat{f}_a}$	$e_{\hat{J}_a}$	$e_{J_a^*}$
a	-0.2525	0.806	0.362	1.011	0.203	0.783	0.768
b	0.2525	0.569	1.164	0.295	-0.273	0.600	0.474
c	0.04366	-0.308	1.093	-0.954	-0.648	-0.236	-0.161
d	0.03207	-0.379	0.635	-0.846	-0.469	-0.586	-0.534
δ	0.00101	0.000	0.000	0.000	0.000	0.000	-0.000
e	0.01254	-0.054	-1.249	0.497	0.551	-0.227	-0.239
ϵ	0.00101	0.000	0.000	0.000	0.000	0.000	-0.000
G	1.616	-0.252	0.601	0.812	0.066	-0.117	-0.080
k	$477,344$	1.000	1.000	1.000	0.000	1.000	0.996
κ	1.01	0.014	-0.201	0.114	0.099	0.749	0.743
μ	0.23014	-0.700	-1.896	-0.150	0.555	-0.762	-0.613
ω	-0.505	0.080	-1.106	0.626	0.546	1.181	1.158
r	0.0404	0.013	-0.184	0.104	0.091	0.003	-1.208
z	0.606	0.016	1.088	-0.478	-0.494	0.071	0.071

Figure 3.5: Steady state value $\hat{A}_a$ as a function of G for per gram social costs \$ 50 (thin curve), \$ 100, and \$ 200 (thick curve), respectively

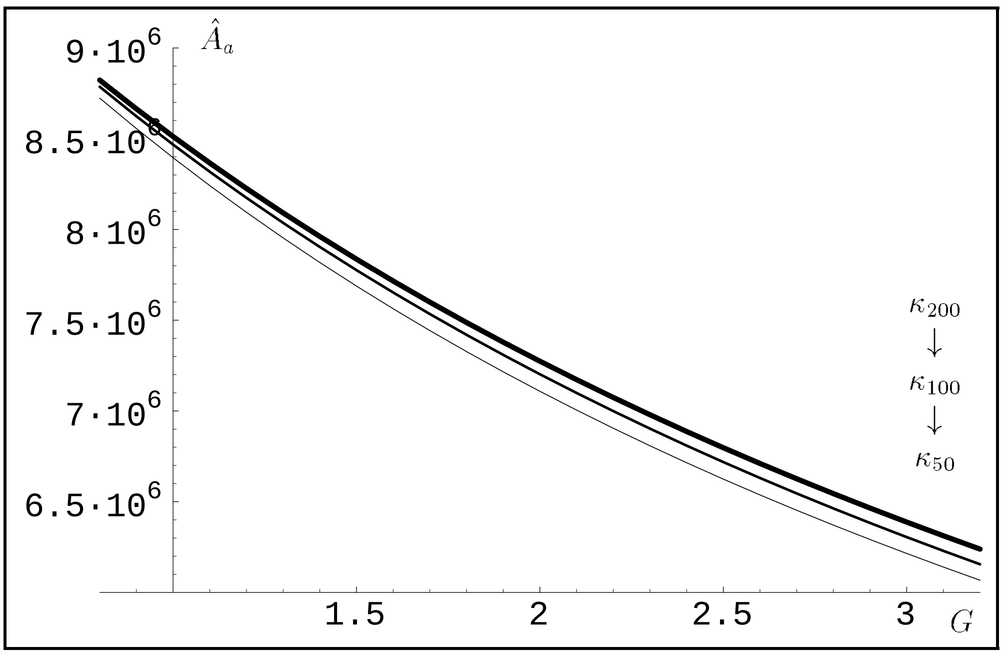


Figure 3.6: Steady state fraction $\hat{f}_a$ as a function of G for per gram social costs \$ 50 (thin curve), \$ 100, and \$ 200 (thick curve), respectively; the vertical line at $G = 1.6$ indicates the base value of G

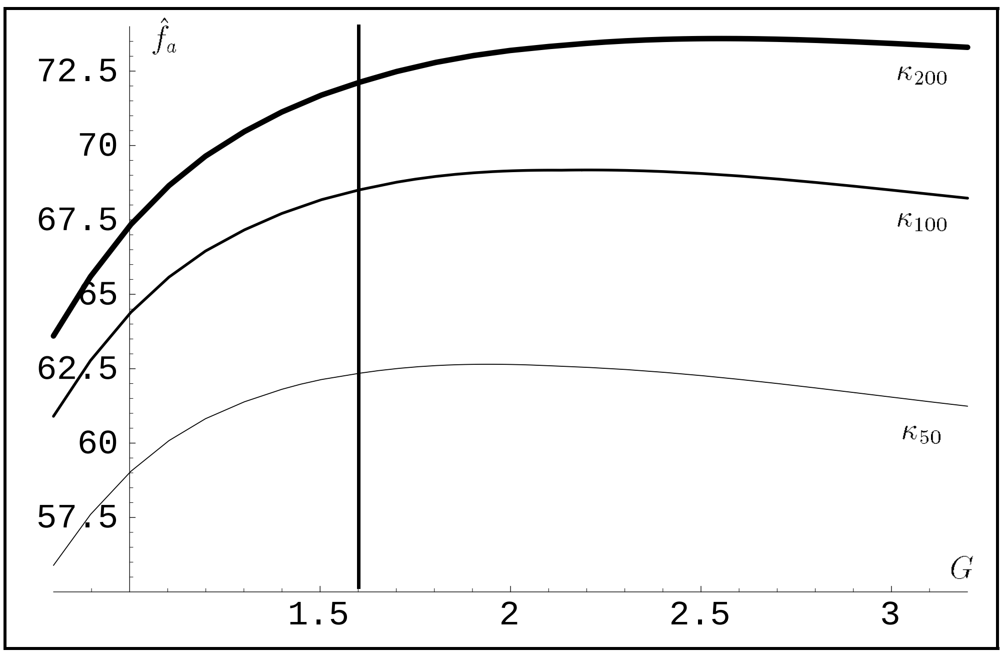
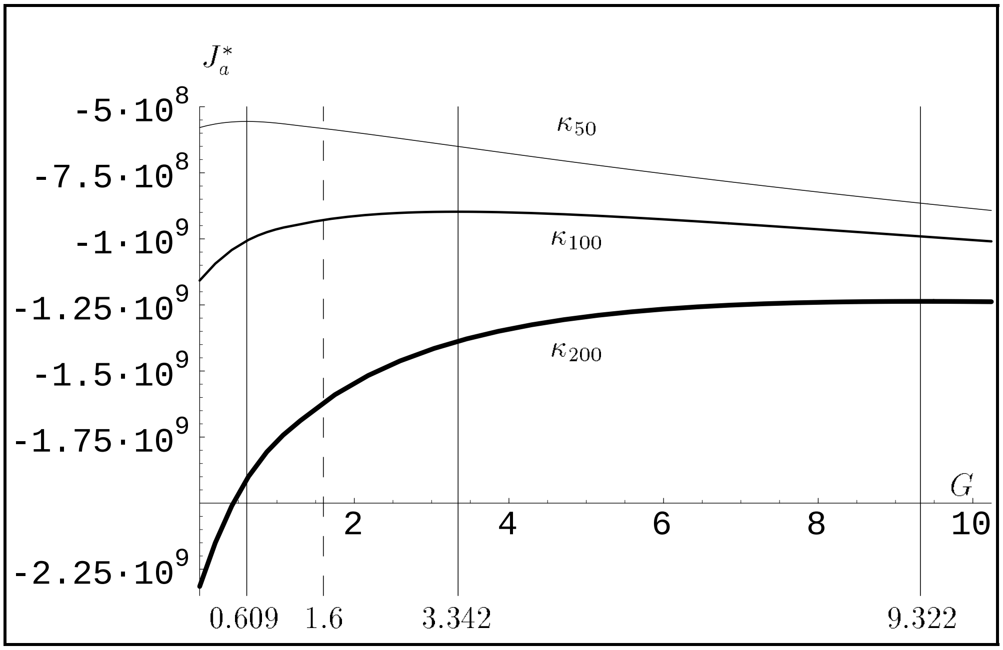


Figure 3.7: Optimal utility functional value J_a^* as a function of G for per gram social costs \$ 50 (thin curve), \$ 100, and \$ 200 (thick curve), respectively; the vertical lines at $G = 0.609$, $G = 3.342$, and $G = 9.322$ indicate those sizes of the per user budgets, which yield the highest utility functional values for the corresponding per gram costs \$ 50, \$ 100, and \$ 200, respectively; the dashed vertical line at $G = 1.6$ represents the base value of G



J_a^* (Fig. 3.7 for a "low" initial state, $A_0 = 100,000$; Fig. 3.8 for a "high" initial state, $A_0 = 6,500,000$), respectively, for per gram costs of \$ 50 (half of the base value; thin curves), \$ 100 (base value), and \$ 200 (double the base value; thick curves) as a function of G, the amount of drug control spending per user.

Fig. 3.5 shows that increasing the per gram costs causes a negligible increase in the steady state value $\hat{A}_a$, whereas $\hat{A}_a$ decreases considerably when the budget per user that the government is willing to spend increases. This raises the question of whether it is possible at all to "eradicate" illicit drug use. I.e. will $\hat{A}_a$ become zero if G becomes high enough? Fig. 3.9 suggests that this does not seem to be the case, at least not for reasonable

Figure 3.8: Optimal utility functional value $J_a^*(A_0 = 6,500,000)$ as a function of G for per gram social costs \$ 50 (thin curve), \$ 100, and \$ 200 (thick curve), respectively; the vertical lines at $G = 0.541$, $G = 2.879$, and $G = 8.167$ indicate those sizes of the per user budgets, which yield the highest utility functional values for the corresponding per gram costs \$ 50, \$ 100, and \$ 200, respectively; the dashed vertical line at $G = 1.6$ represents the base value of G

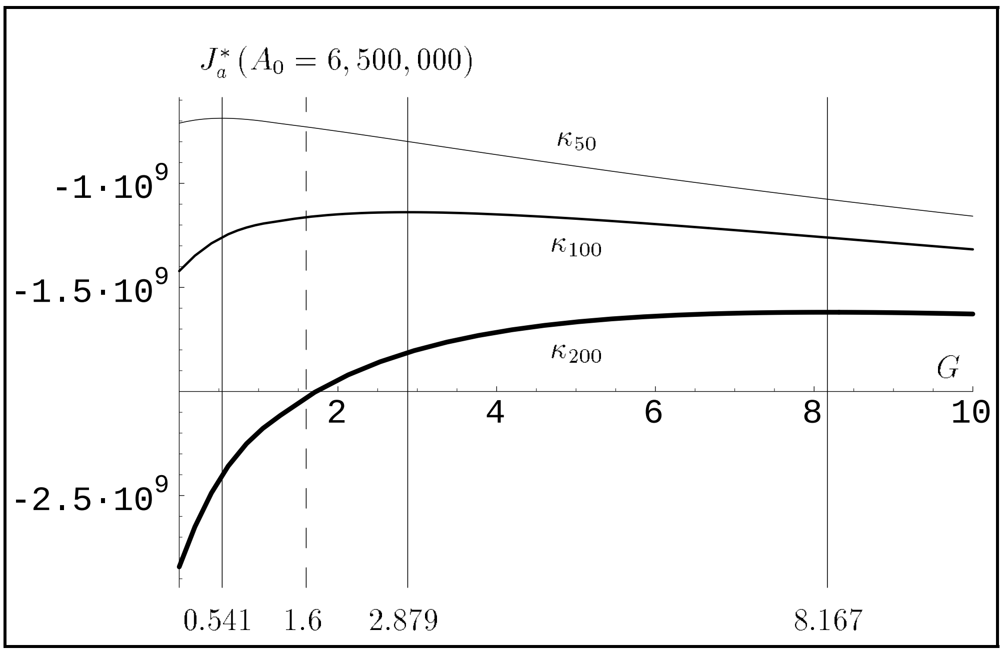
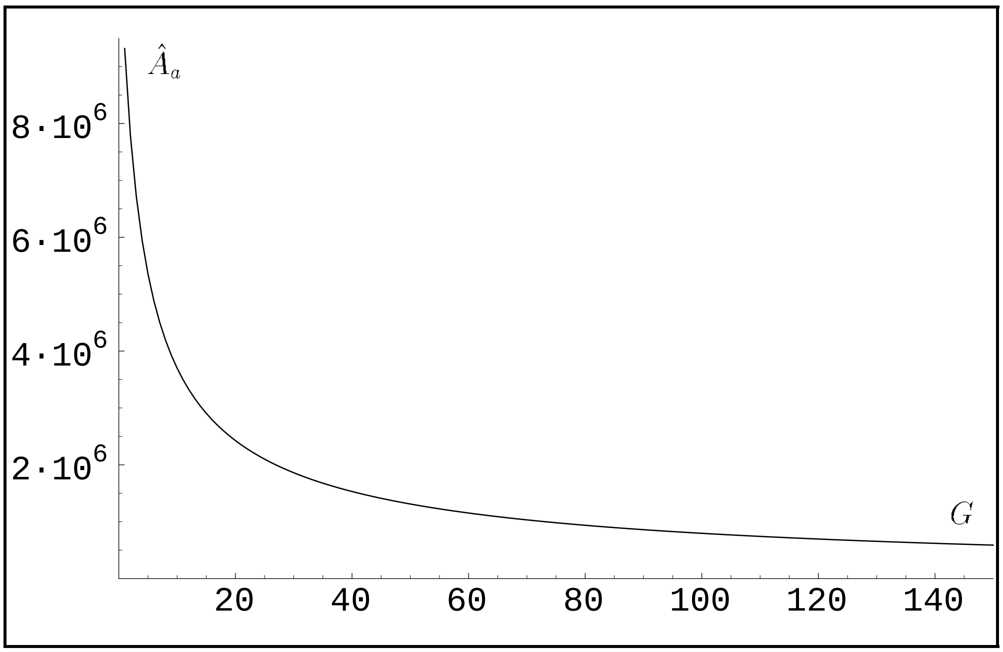


Figure 3.9: $\hat{A}_a$ as a function of G; even for considerably higher values of G (nearly 100 times as high as the base value), it is not possible to eradicate illicit drug use



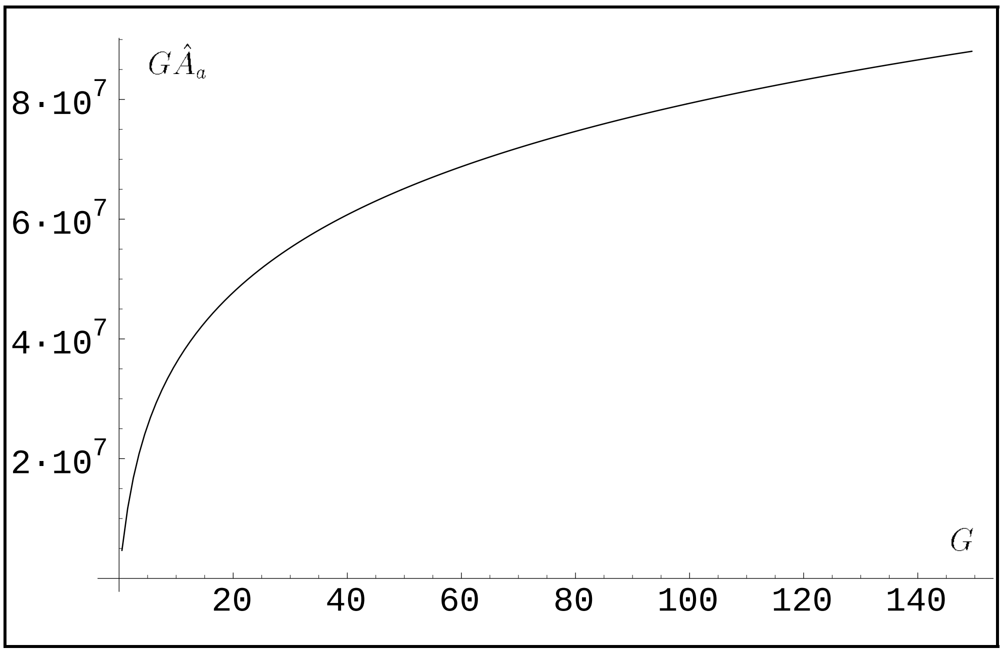
values of G. We will pose the same question in a different context once more in Chapter 6 – and there it will lead to a different answer.

Is it possible that higher drug control spending per user might reduce total drug control spending? That question is similar to the one raised by the so-called Laffer curve (see, e.g., Varian, 1996). The Laffer curve shows government tax revenues as a function of the tax rate. If the tax rate is zero, tax receipts are zero. As the tax rate increases, tax revenues increase, but not quite proportionately because as taxes gobble up a larger and larger fraction of marginal income, people will not have an incentive to work as hard. In the extreme, if the tax rate were 100 %, presumably no one would work because they would not be able to keep the benefits of their labor, so income and, thus, tax revenues would be zero. Some hypothesized that in 1980 U.S. tax rates were so high that the country was on the right side of the curve and that cutting the tax rate would increase, not decrease, tax revenues. That turned out not to be the case, as the Reagan tax cut increased the budget deficit, but one can ask a similar question about the level of drug control spending (GA) as a function of drug control spending per user (G). Obviously spending is zero if spending per user is zero. Presumably as G increases, the number of users decreases, so spending in equilibrium ($G\hat{A}$) is concave in G. Perhaps it even bends down because if G is high enough, the equilibrium number of users might be very low. If so, then there ought to be a level of spending such that the government could save money by increasing spending per user. Cynics might think that the U.S. is near the maximum of the GA curve so the U.S. could save money by either cutting G or increasing G.

That would be a wonderfully ironic story, but Fig. 3.10 suggests that the U.S. is very far to the left of the peak. Currently G in the U.S. is about 1.6, and the $G\hat{A}$ curve is still rising for G above 150. Strictly speaking Fig. 3.10 shows spending in steady state, whereas the relevant plot is the total discounted spending (NPV over time of GA) as a function of G. The NPV of GA depends on the initial conditions, but cocaine consumption has been fairly stable in the U.S. for over five years, suggesting that the number of users is fairly close to the equilibrium now. Hence, the curve of NPV of GA vs. G starting today probably looks like the curve in Fig. 3.10 at least in broad terms.

An interesting result follows from Fig. 3.6. The higher the per gram costs are estimated to be, the higher will be enforcement's share of the budget in the steady state. On the other hand, for any per gram cost there seems to

Figure 3.10: $G\hat{A}_a$ as a function of G; even for considerably higher values of G (nearly 100 times as high as the base value), it is not possible to reduce the total budget spent at the steady state, $G\hat{A}_a$



exist one value of G where enforcement's share reaches a maximum, i.e. for values of G smaller or bigger than that special value the government will rely more on treatment. For the base case parameter values, the willingness to spend (G) is quite close to the value that maximizes the share of the budget that is optimally allocated to enforcement.

Figs. 3.7 ($A_0 = 100,000$) and 3.8 ($A_0 = 6,500,000$) illustrate the dependences of the optimal utility functional value J_a^* both on different values of the per gram costs and on the budget per user. First, the total social loss is bigger the higher the costs per gram consumed are, which makes sense. Secondly, it turns out that for any given value \$ *pgc* of the per gram costs, there seems to exist one value of G, G_{pgc}^*, for which the utility functional is maximized, where

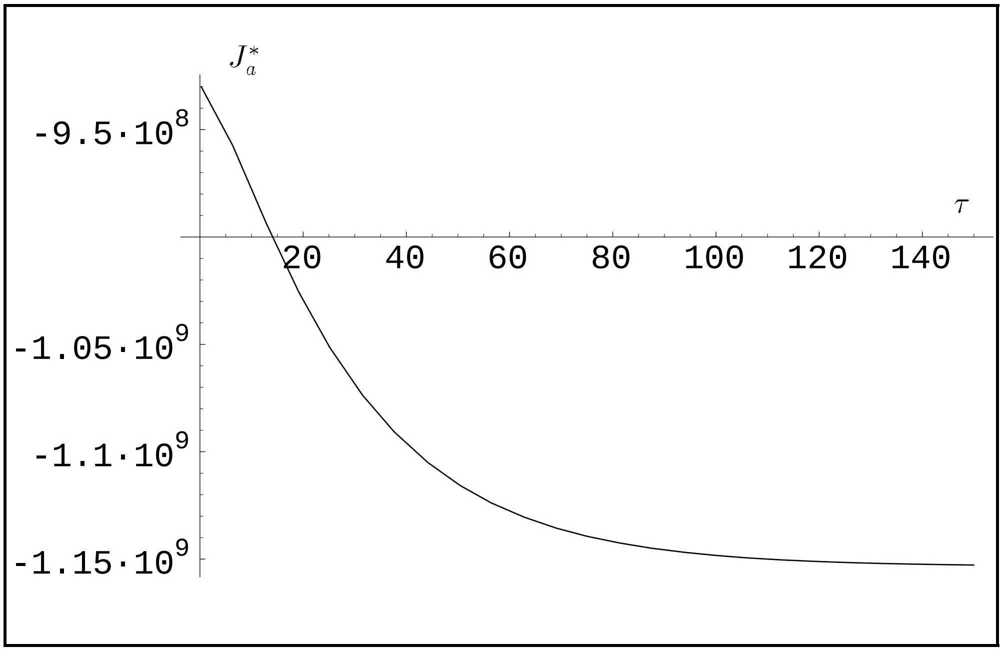
$$G_i^* < G_j^* \text{ if } i < j.$$

The consequence is that the slope of the optimal utility functional value curves at any given level of G also depends on the size of the per gram costs as follows. Suppose G is at 1.6 (base case). Then, if the social costs are low (\$ 50; thin curve), the government would be better off by cutting the size of the drug control budget per user, whereas for high per gram costs (\$ 200; thick curve) it would pay to increase the budget. Also, for the base case value of \$ 100 there is a slight advantage to having a larger G. However, looking at Figs. 3.7 and 3.8 we see that there probably exists a value of the per gram costs (somewhere between \$ 50 and \$ 100) where the government could neither win nor lose by increasing or decreasing the budget, at least in a wide range around the base value of G.

Further, comparing the utility functional value curves for different (low and high, respectively) initial states A_0 shows first that the results do not change qualitatively, if A_0 is changed. However, the optimal values of the per user budgets G decrease if A_0 increases, while the overall damage increases. As starting the control problem with a higher value of the initial number of users can also be interpreted as delaying the beginning of the control, we next want to investigate the question of delays in the controls.

Realistically, it may take time from the beginning of a drug epidemic until the government can start to impose controls for that specific drug problem. The reasons for this lag could be manifold: it takes time for people to become aware of a drug epidemic, the government may not want to intervene when the problem is small, or – even if the government decides to do something

Figure 3.11: Optimal utility functional value J_a^* as a function of τ, where τ denotes the time when the government starts to control



– it may take time until the whole bureaucracy is established, etc. For this reason, in Fig. 3.11 we plot the optimal utility functional value as a function of τ, with τ denoting how long it is after the point when there are 100,000 users that the government starts to control. To accomplish this, we run the uncontrolled model for τ years. This results in a certain number of users at that time, $A(\tau)$. Starting with that initial condition, we then run the optimally controlled allocation problem (with the standard infinite planning horizon assumption). Then, the optimal utility functional values from both problems are added by properly discounting the value of the second stage problem. As we can see from Fig. 3.11, one could save up to 25 % of the social costs by starting to control early in the epidemic, and the costs of delaying an additional year diminish the longer one waits.

The reduction in the value of the objective function is nearly linear over the first 20 years of delay, and actual delays are likely to be no more than 20 years. So it makes sense to think of a cost per year of delay, and that figure is about \$ 5.5 billion per year (net present value of future costs).

Suppose one is not currently in a drug epidemic. (That might be a fair description of the cocaine situation in Austria at present.) Suppose further that the expected time until the next epidemic starts is 25 years. Then (ignoring the time value of money) one should be willing to spend up to

$$\frac{\$\ 5.5 \text{ billion}}{25} = \$\ 220 \text{ million}$$

per year on a monitoring system that would let you detect the next epidemic's start a year earlier than you could without the monitoring system. That is substantially more than the U.S. has ever spent monitoring drug problems. Of course since Austria is about $\frac{1}{33}$ as large as the U.S., the savings in Austria would be $\frac{1}{33}$ as great as in Fig. 3.11. Thus Austria should be willing to spend up to \$ 6.67 million per year for a monitoring system that could detect the outbreak of such an epidemic a year earlier than it would otherwise be detected.

3.5 The Unrestricted Control Problem

The current value Hamiltonian for (2.8) with (2.7) subject to (2.6) is written

$$H = -\kappa A p^{\omega} - u - v + \lambda \left(k p^{a} - c\beta^{z} A - \mu p^{b} A \right).$$

Here, $H_u = 0$ yields

$$\lambda = \frac{1}{-cz\beta^{z-1}\beta_u A} < 0 \tag{3.11}$$

which implies that

$$H_{uu} = \lambda cz(1-z)\beta^{z-2}{\beta_u}^2 A < 0$$

and

$$H_{vv} = {p_v}^2 \left\{ \kappa\omega(1-\omega)p^{\omega-2}A + \lambda \left[a(a-1)kp^{a-2} - \mu b(b-1)p^{b-2}A \right] \right\} < 0,$$

i.e. H is concave both in u and in v. The concavity of the maximized Hamiltonian with respect to the state variable, however, cannot be guaranteed, so again the usual sufficiency conditions are not satisfied.

Table 3.7: Steady state values $\hat{A}_f$, $\hat{u}_f$, and $\hat{v}_f$, enforcement's share at steady state, $\hat{f}_f$, and optimal utility functional value, J_f^*, in the unrestricted optimal control problem

$\hat{A}_f$	$\hat{u}_f$	$\hat{v}_f$	$\hat{f}_f$	J_f^*
$6,395,040$	$5,779,559$	$12,665,715$	0.687	$-886,110,702$

From $H_v = 0$ we get

$$\lambda = \frac{1 + \kappa\omega p^{\omega-1} p_v A}{p_v\left(akp^{a-1} - \mu bp^{b-1}A\right)} =: \frac{Z(\lambda)}{N(\lambda)}. \tag{3.12}$$

Equating (3.11) and (3.12) allows to express u as a function of A and v:

$$u = \left[\frac{cz{\beta_u}^z A\left(1 + \kappa\omega p^{\omega-1} p_v A\right)}{p_v\left(\mu bp^{b-1}A - akp^{a-1}\right)}\right]^{\frac{1}{1-z}}$$

(note that $\beta_u = \frac{1}{A+\delta}$ does not depend on u).

As in the allocation problem we derive

$$\dot{v} = \frac{N(\lambda)\left[Z(\lambda)\left(r - H_A^{(2)}\right) - N(\lambda)H_A^{(1)}\right] - Z\left(\lambda_A\right)\dot{A}}{Z\left(\lambda_v\right)} \tag{3.13}$$

with $N(\lambda)$ and $Z(\lambda)$ from (3.12) and the other functions as given in Appendix A.5.

Again, we use the base parameter values from Table 3.1 to compute the steady state values for this problem by setting $\dot{A}$ and $\dot{v}$ equal to zero (grey curves in Fig. 3.12). As was the case for the allocation problem, the intersection of the isoclines is a saddle point equilibrium $\left(\hat{A}_f, \hat{v}_f\right)$ (see Table 3.7), and the two stable manifolds (thick black curves) yield the optimal trajectories. In contrast to the allocation problem above, here we don't have a budget constraint (2.11) that relates the two controls u and v in a "simple" way, so it makes sense to have the additional Fig. 3.13, which gives the stable manifolds of the saddle point equilibrium not only in the A-v-plane (black) but also in the A-u-plane (grey). It is striking that these two curves are fairly

Figure 3.12: Phase portrait in the A-v-plane for the unrestricted optimal control problem using the base parameter values from Table 3.1; the grey curves represent the $\dot{A} = 0$ and $\dot{v} = 0$ isoclines, respectively, the thick black curves are the stable manifolds of the saddle point equilibrium $\left(\hat{A}_f, \hat{v}_f\right)$ (intersection of the isoclines), the black curves are the unstable manifolds of the steady state, and the thin black curves indicate the flow

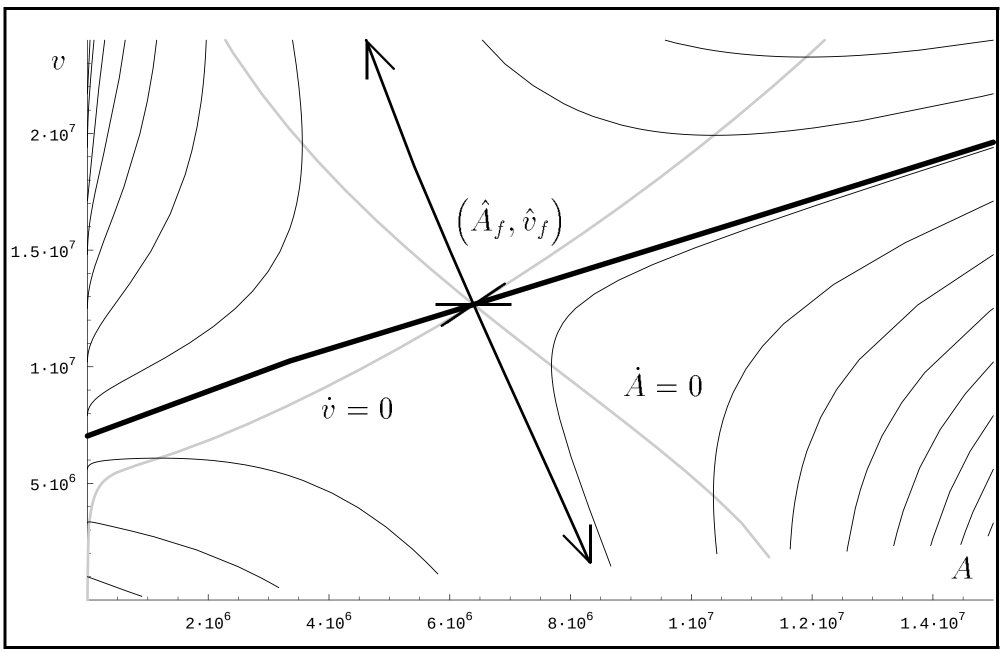


Figure 3.13: The optimal trajectories u (grey) and v (black) as functions of A; the vertical line represents the steady state value $\hat{A}_f$

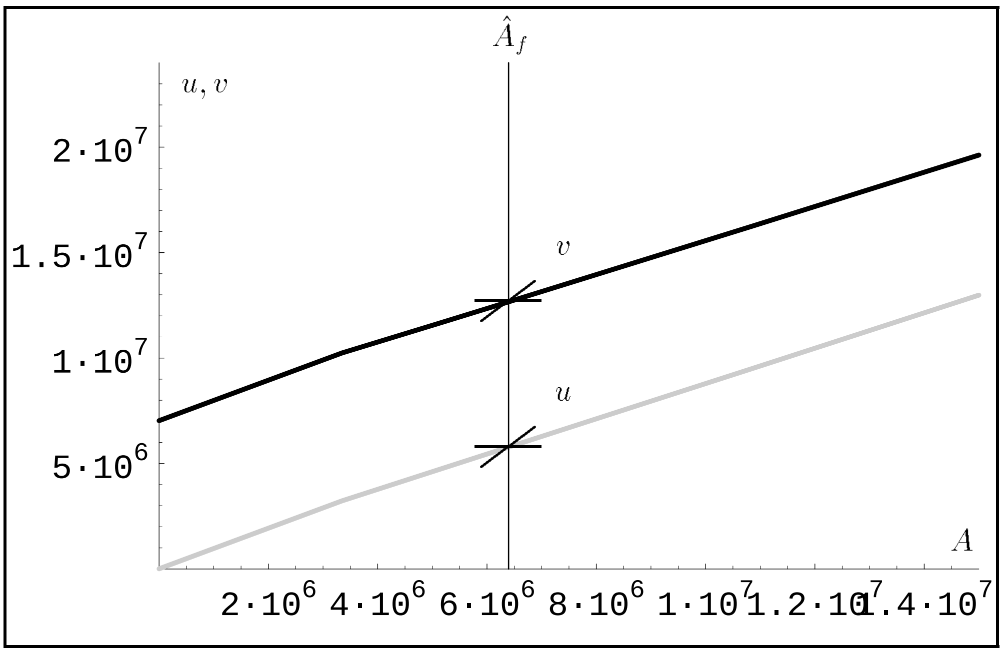
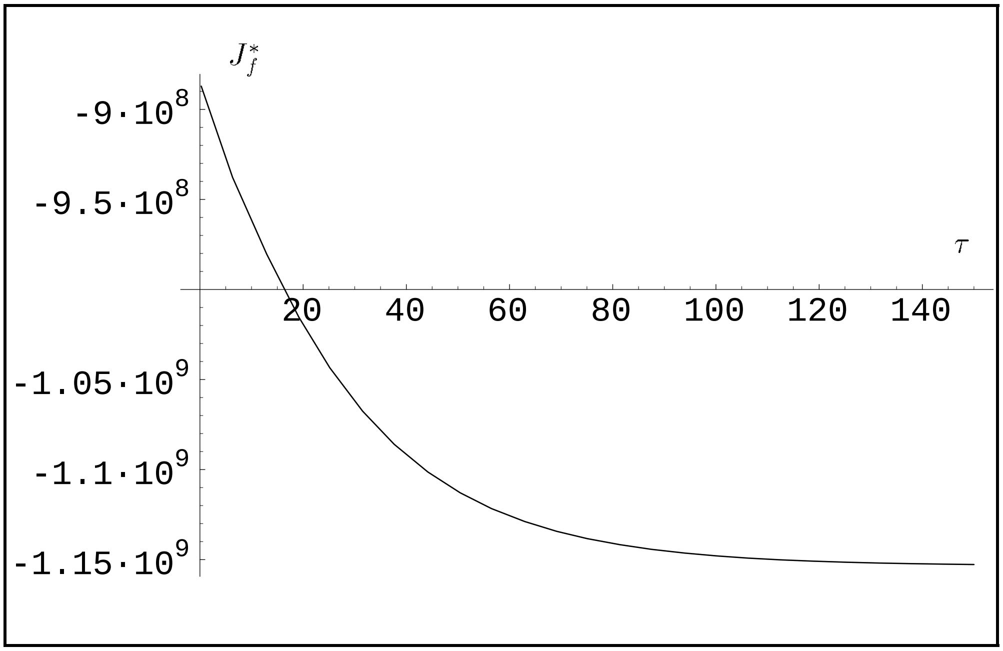


Table 3.8: Percent changes $e_{\hat{A}_f}$, $e_{\hat{u}_f}$, $e_{\hat{v}_f}$, $e_{\hat{f}_f}$, $e_{\hat{J}_f}$, and $e_{J_f^*}$ of $\hat{A}_f$, $\hat{u}_f$, $\hat{v}_f$, $\hat{f}_f$, $\hat{J}_f$, and J_f^* with respect to 1 % increases of the base parameters (”elasticities”)

Para- meter	1.01· · (Base Value)	$e_{\hat{A}_f}$	$e_{\hat{u}_f}$	$e_{\hat{v}_f}$	$e_{\hat{f}_f}$	$e_{\hat{J}_f}$	$e_{J_f^*}$
a	-0.2525	0.559	0.716	1.282	0.176	0.687	0.671
b	0.2525	0.266	1.402	0.674	-0.226	0.472	0.400
c	0.04366	-0.619	1.377	-0.464	-0.576	-0.342	-0.209
d	0.03207	0.269	-0.751	-1.579	-0.263	-0.305	-0.347
δ	0.00101	0.000	0.000	0.000	0.000	0.000	0.000
e	0.01254	-0.308	-0.598	0.569	0.365	-0.343	-0.321
ϵ	0.00101	0.000	0.000	0.000	0.000	0.000	0.000
k	$477,344$	1.000	1.000	1.000	0.000	1.000	0.993
κ	1.01	-0.677	1.227	1.292	0.015	0.448	0.550
μ	0.23014	-0.534	-1.926	-0.372	0.491	-0.687	-0.569
ω	-0.505	-1.209	1.758	2.684	0.283	0.595	0.776
r	0.0404	0.155	-0.434	-0.164	0.085	0.033	-1.200
z	0.606	-0.439	2.044	-0.155	-0.685	-0.080	0.008

Figure 3.14: Optimal utility functional value J_f^* as a function of τ, where τ denotes the time when the government starts to control



parallel, both are increasing in absolute values, and treatment's share in the whole budget is increasing in A.

Table 3.8 gives the percent changes of the steady state values $\hat{A}_f$, $\hat{u}_f$, and $\hat{v}_f$, as well as enforcement's share at steady state, $\hat{f}_f$, and the utility functional values $\hat{J}_f$ (at steady state) and J_f^* (overall) with respect to 1 % changes of the base parameter values. As in the allocation problem, increasing δ or ϵ by 1 % has no effect, and $\hat{A}_f$, $\hat{u}_f$, $\hat{v}_f$, and $\hat{J}_f$ increase by exactly 1 % if k is increased by that amount (J_f^* increases by 0.993 %). In contrast to the allocation problem it is the short run elasticity of demand by current users, ω, that affects the steady state values most seriously (if we take again the sum of the absolute values as a measure). Again, changing r has the smallest impact, and the changes of $\hat{u}_f$ are strongest.

How the time when the government starts to control influences the optimal utility functional value can be seen in Fig. 3.14. As in the previous section with the allocation problem it makes a big difference if the control starts early in the epidemic or at later stages (differences up to 30 % possible). Again the costs per year of delay are roughly linear for $\tau \leq 20$ years, but the consequences of such delays are about 20 % greater or roughly \$ 6.6 billion per year.

Chapter 4

Comparing the Results of the Four Different Models

Combining Tables 3.2, 3.5, and 3.7, we get Table 4.1 which shows the steady state values $\hat{A}$, $\hat{u}$, and $\hat{v}$, enforcement's share of the control budget in steady state $\hat{f}$, and the optimal utility functional values J^*, for all four problems considered in this paper. We see that

$$\hat{A}_n > \hat{A}_c > \hat{A}_a > \hat{A}_f$$

and

$$J_n^* < J_c^* < J_a^* < J_f^*,$$

respectively, which are the sequences that one would expect.

To help visualize how much these values differ across the four problems, Figs. 4.1 and 4.2 illustrate the percent values of $\hat{A}$ and J^*, respectively,

Table 4.1: Steady state values, enforcement's share at the steady state, and optimal utility functional values for all four models

Control	$\hat{A}$	$\hat{u}$	$\hat{v}$	$\hat{f}$	J^*
n	11,640,259	0	0	–	– 1,157,597,439
c	7,718,898	3,421,016	8,929,221	0.723	– 930,532,431
a	7,652,824	3,857,023	8,387,495	0.685	– 912,889,670
f	6,395,040	5,779,559	12,665,715	0.687	– 886,110,702

Figure 4.1: Percent values of $\hat{A}$ for all models relative to the steady state value in the uncontrolled model

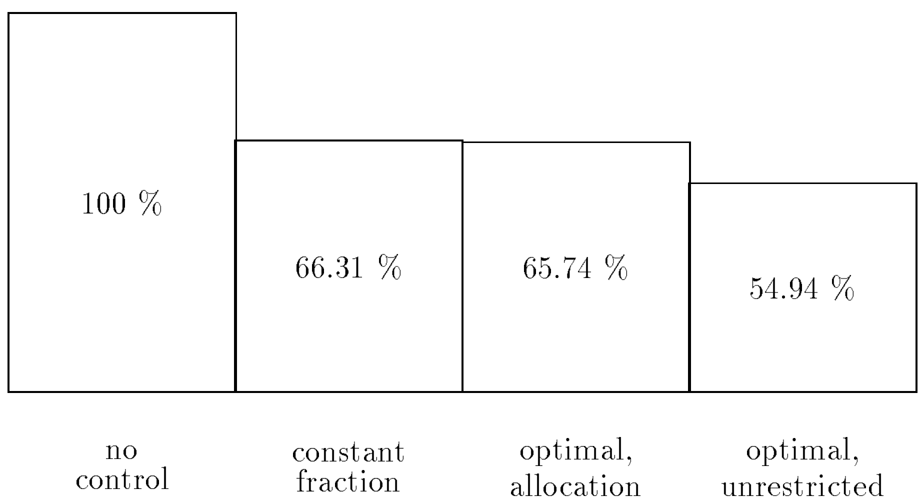


Figure 4.2: Percent values of J^* for all models relative to the utility functional value in the uncontrolled model

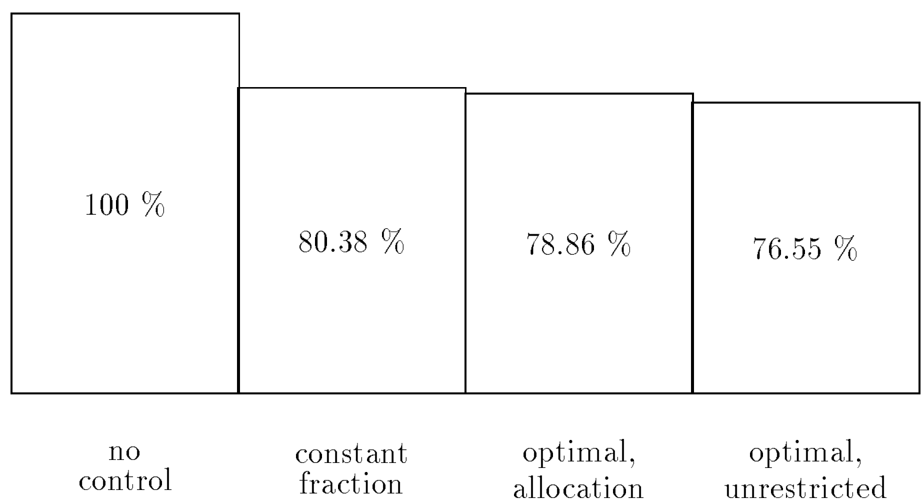
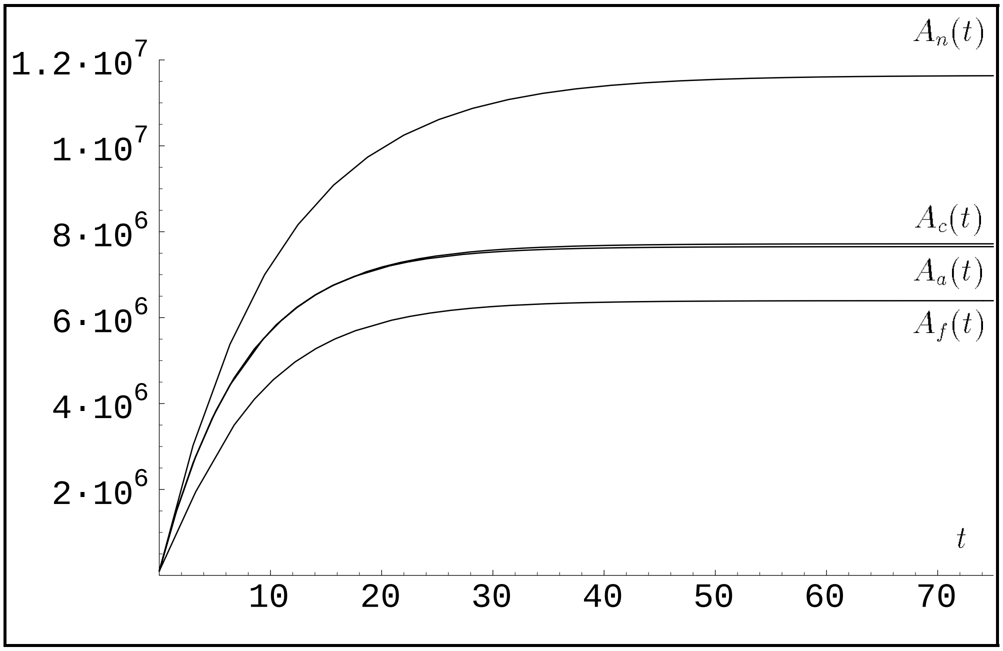


Figure 4.3: Time paths $A_n(t)$ (top), $A_c(t)$ (middle, top), $A_a(t)$ (middle, bottom), and $A_f(t)$ (bottom) for $t \in [0, 75]$, starting with $A_0 = 100,000$

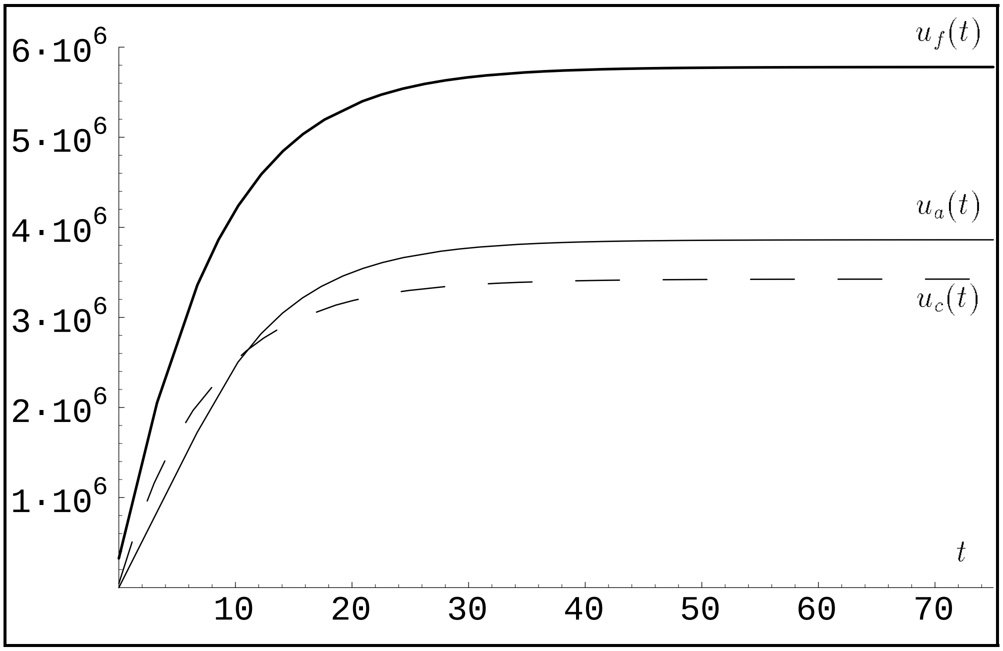


relative to the (100 %) reference point of no control. We see that the controls in the two allocation problems cut down the steady state value $\hat{A}$ by about one third, whereas the optimal unrestricted control reduces the number of users in steady state by some additional 10 %. A similar result follows from Fig. 4.2 for the values J^*. The total social loss is reduced by about 20 % if any of the three control mechanisms is imposed; however, it is interesting to see that the "dumb" constant fraction control yields an optimal utility functional value which is only a few percent worse than the best (i.e., the unrestricted) control.

Up to now we have concentrated mainly on the steady state values of the four problems. However, the problems considered in this paper are dynamic, so it is interesting to observe how the state and control variables change with time.

Fig. 4.3 gives the time paths of the state variable $A(t)$, $t \in [0, 75]$, for all four problems. All four curves look qualitatively the same, but – as already

Figure 4.4: Time paths of treatment spending $u_c(t)$ (dashed), $u_a(t)$ (thin), and $u_f(t)$ (thick) for $t \in [0, 75]$, starting with $A_0 = 100,000$



noted – the steady states approached are different. Note that the number of addicts grows rapidly early in the epidemics but changes more slowly as the steady state is approached. It is also interesting to see that – apart from the unrestricted control case – the time paths stay very close to one another during the first few years before they diverge.

The dynamics of the controls $u(t)$ and $v(t)$, $t \in [0, 75]$, are illustrated in Figs. 4.4 and 4.5, respectively. The most striking result that comes from these figures is that the temporal evolution of the controls in the two allocation problems is very similar, which seems to be the reason why all the other results are so similar for these two models.

However, there is an enormous gap between the controls in the allocation problems and those in the unrestricted control problem. Treatment spending $u_f(t)$ seems to be roughly some multiple of $u_a(t)$, but enforcement spending $v_f(t)$ seems to be $v_a(t)$ plus some (big) constant. In other words, the total budget spent in the unrestricted control case is significantly higher early in

Figure 4.5: Time paths of enforcement spending $v_c(t)$ (dashed), $v_a(t)$ (thin), and $v_f(t)$ (thick) for $t \in [0, 75]$, starting with $A_0 = 100,000$

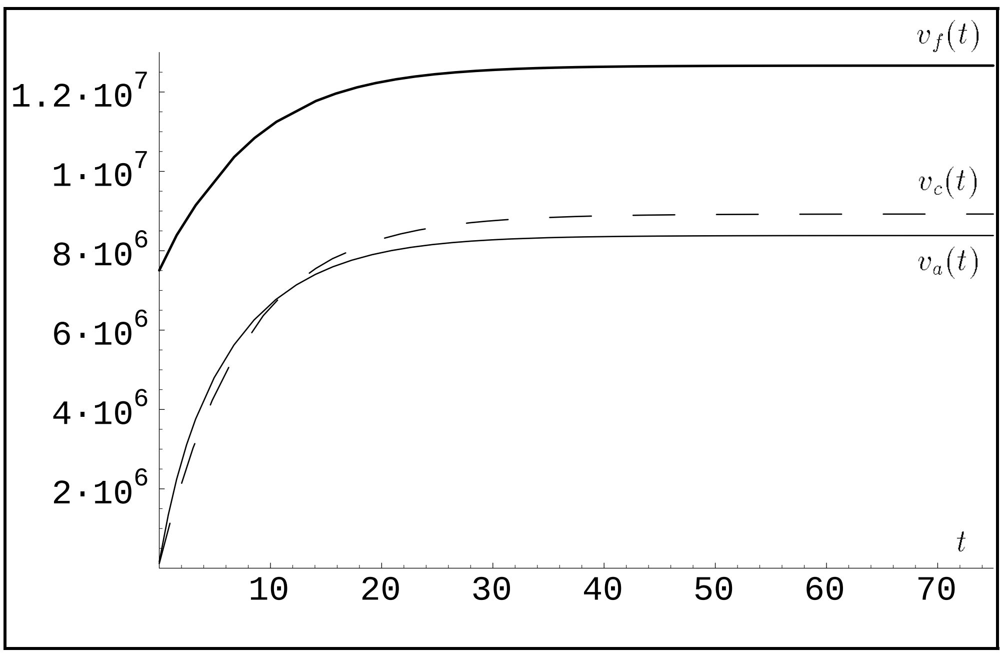
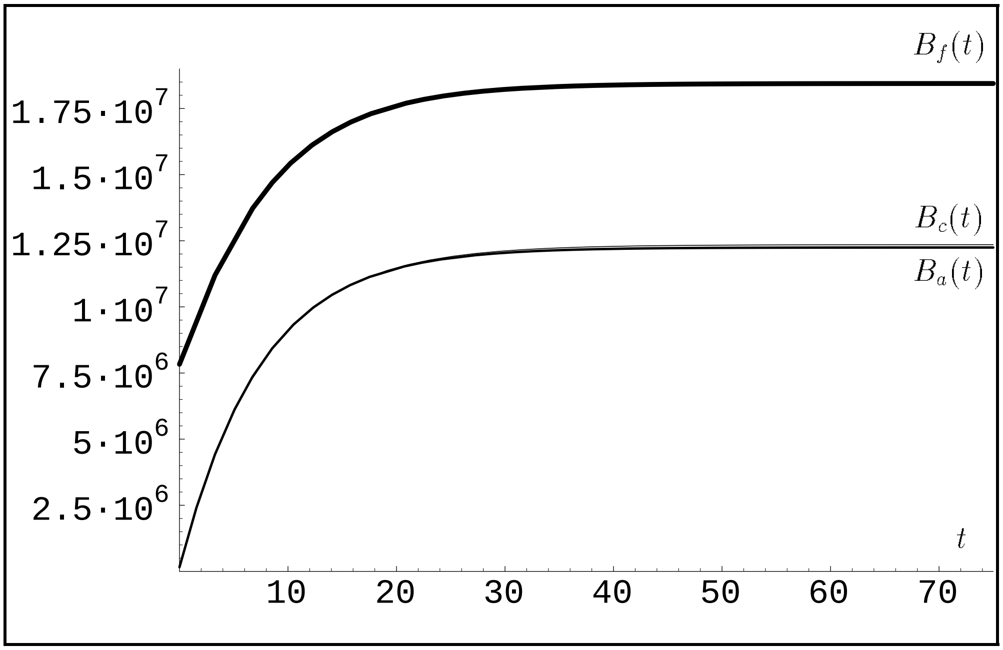


Figure 4.6: Time paths of total drug control spending $B_c(t)$ (thin), $B_a(t)$ (regular), and $B_f(t)$ (thick) for $t \in [0, 75]$, starting with $A_0 = 100,000$; $B_c(t)$ and $B_a(t)$ are nearly "one curve"



the epidemics than it is for the allocation problem. The government chooses a high level of enforcement (about two-thirds that of the steady state value) from the beginning if the size of the budget is unconstrained. Looking back to the utility functional values in Table 4.1 (see also Fig. 4.2), however, we find out that what the government receives from all this additional effort is rather modest. For a dynamic comparison, the temporal evolutions of both the total budgets for the three control problems (constant fraction, allocation and unrestricted) and the social costs due to consumption for all four problems are given in Figs. 4.6 and 4.7, respectively. From these figures and also Fig. 4.8, which illustrates the budget (in \$ 1,000) that is spent per user in the two control problems, we therefore conclude that it might be difficult for the government to justify these high levels of expenditures, if benefits are almost entirely offset by the greater costs.

To see how the proportions of treatment and enforcement spending in the

Figure 4.7: Time paths of social costs due to consumption $C_n(t)$ (dashed), $C_c(t)$ (thin), $C_a(t)$ (regular), and $C_f(t)$ (thick) for $t \in [0, 75]$, starting with $A_0 = 100,000$, where $C(t) := \kappa p\left(A(t), v(t)\right)^{\omega} A(t)$; $C_c(t)$ and $C_a(t)$ are nearly "one curve"

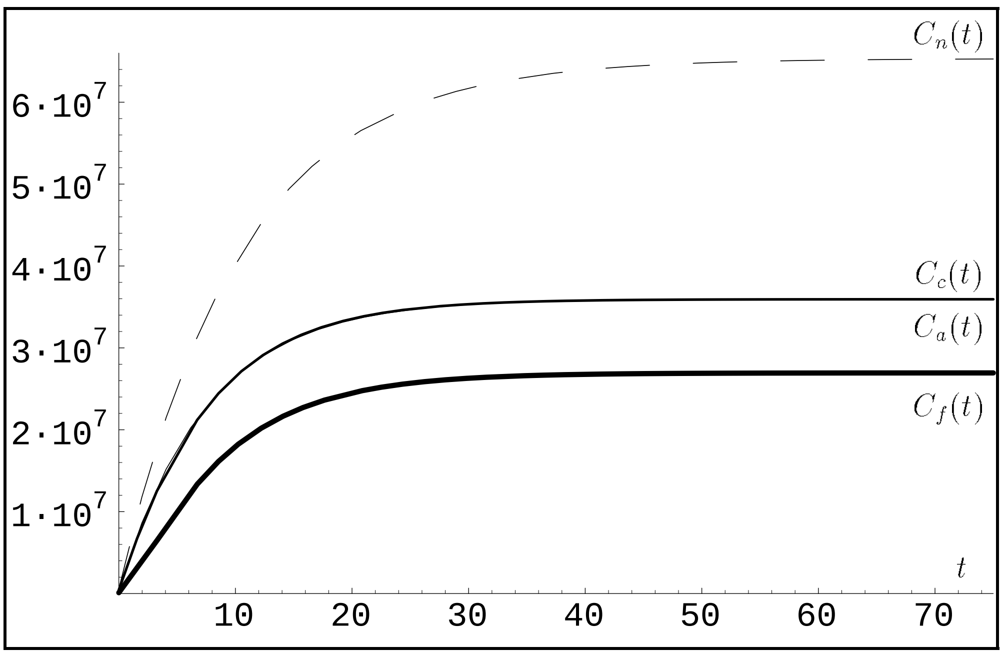


Figure 4.8: Time paths $G_a(t) \equiv G$ (lower curve) and $G_f(t)$ (upper curve) for $t \in [0, 75]$, starting with $A_0 = 100,000$

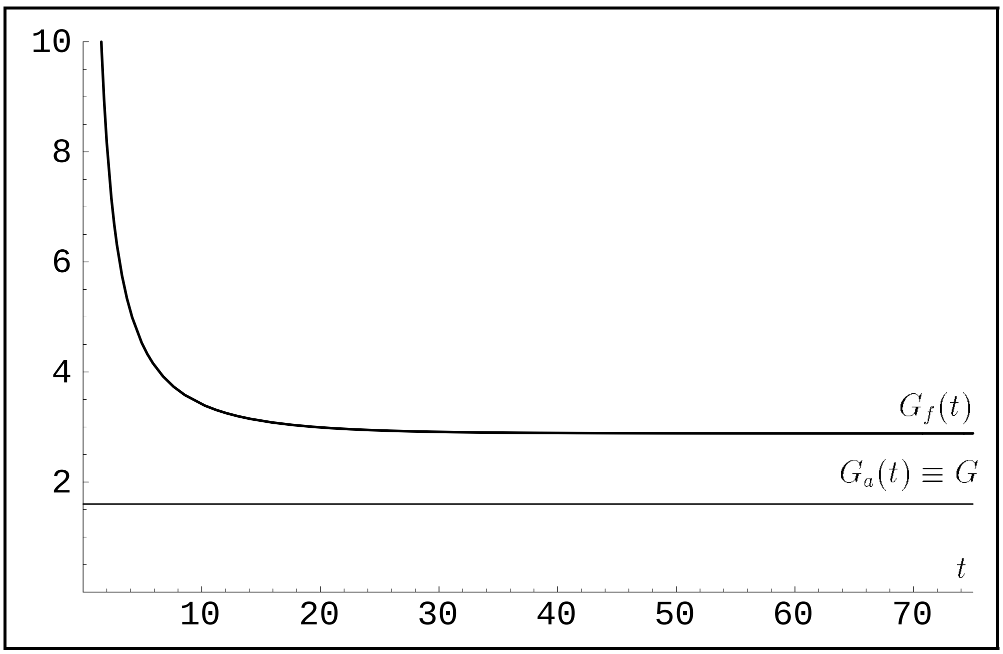
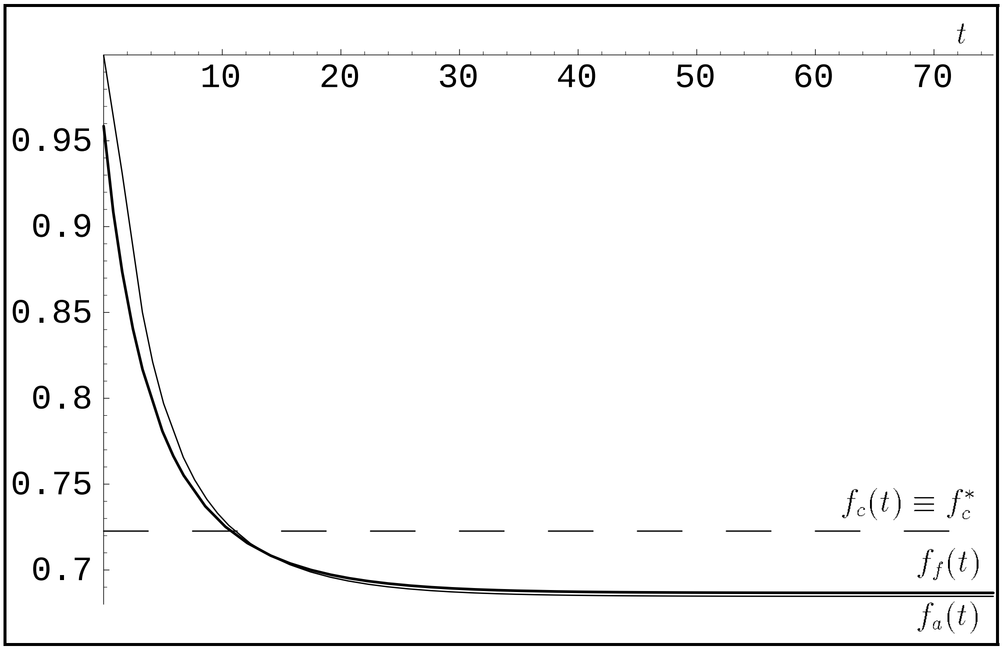


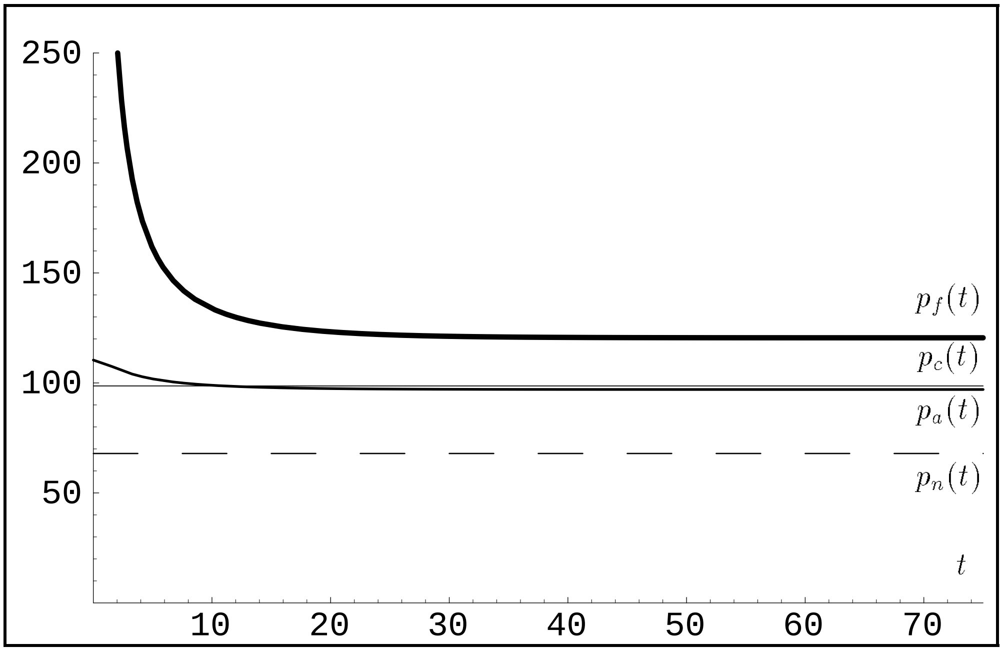
Figure 4.9: Enforcement's share in the whole budget, $f(t) := \frac{v(t)}{u(t)+v(t)}$, for the constant fraction control (dashed), the optimal control in the allocation problem (thin), and the optimal unrestricted control (thick), respectively, for $t \in [0, 75]$, starting with $A_0 = 100,000$



total budget evolve over time, Fig. 4.9 plots enforcement's share of the budget over time along the optimal trajectory. In the optimally controlled allocation problem, the government starts with 100 % enforcement which is gradually decreased to its steady state value. The development of enforcement's share in the unrestricted control problem is similar, but there treatment starts with a share of around 5 %. Finally, in the constant fraction model the government chooses a fixed portion which can be seen as some kind of "average level", in the long run lying above the steady state values of the optimal control problems.

Just because the optimal and constrained models produce similar objective function values, does not mean they produce similar control programs or similar epidemics. For example, we have already seen that the number of users over time is different for different control models. An even more

Figure 4.10: Time paths of prices $p_n(t)$ (dashed), $p_c(t)$ (thin), $p_a(t)$ (regular), and $p_f(t)$ (thick) for $t \in [0, 75]$, starting with $A_0 = 100,000$



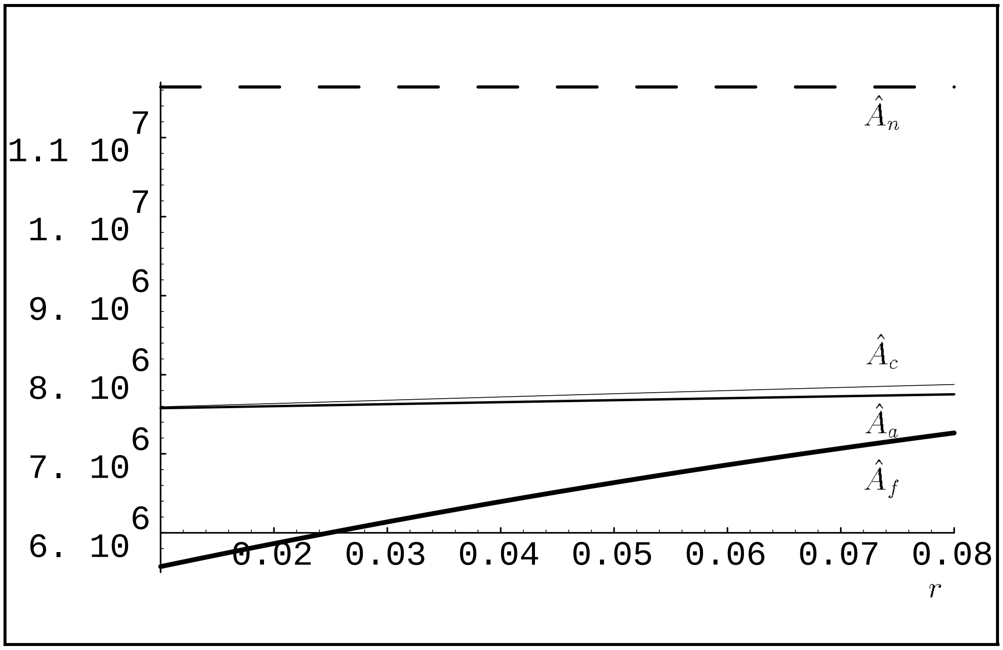
dramatic difference is that the optimal control policy calls for greater drug control spending, enormously greater in percentage terms in the early years.

Since most of that spending is devoted to enforcement, in the early stages of the epidemic, prices are much higher under optimal control than they are with constant spending per user (see Fig. 4.10). That means the optimal policy has much lower consumption in the early years than do these other policies even though differences in the number of users are not so great. However, the benefits of lower consumption are substantially offset by the greater cost of the drug control program.

Also, note that it is optimal to have prices collapse in the early years of the epidemic. That in fact happened for cocaine in the U.S. during the 1980s. Often that precipitous price decline is thought of as a disaster, and it may have been a disaster, but it is also possible that it was the consequence of an optimal policy.

One general insight from these plots vs. time for the different models

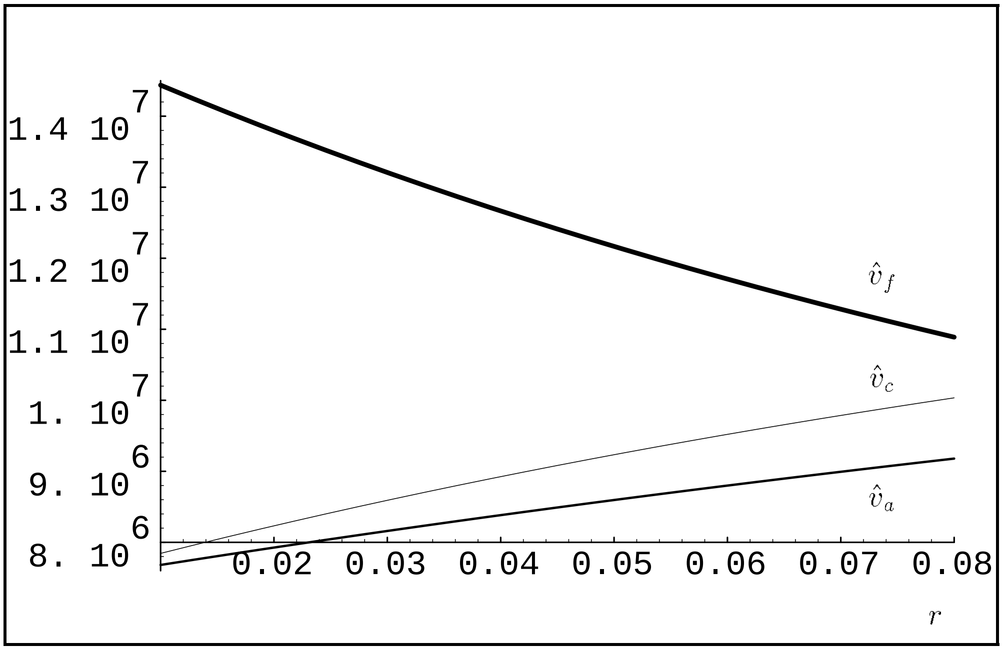
Figure 4.11: $\hat{A}_n$ (dashed), $\hat{A}_c$ (thin), $\hat{A}_a$, and $\hat{A}_f$ (thick) as functions of r



is that, with the exception of the price trajectory, they are generally very similar. They differ in magnitude but have the same basic shape. Since the magnitude of the epidemic is directly influenced by k and k cannot be observed directly, this has an interesting implication. Even towards the end of an epidemic, it will not be easy to look back at the time trajectories and judge whether the government did an optimal or an awful job of controlling the epidemic. The trajectories of a poorly controlled epidemic and an optimally controlled epidemic with a larger k would be nearly indistinguishable.

In Fig. 4.3 we saw that although at early stages of a drug epidemic the time paths of the numbers of users stay rather close in all four models, the differences tend to become fairly big later on. As A can be viewed as a measure of the size of the drug problem, one could suppose that a government that is less myopic will profit more by imposing controls on it. The curves in Fig. 4.11 illustrate the dependence of all four problems' steady state values of the numbers of users on the discount rate r. Obviously, $\hat{A}_f$ of the unrestricted control problem reacts most strongly to changes in r, whereas

Figure 4.12: $\hat{v}_c$ (thin), $\hat{v}_a$ (regular), and $\hat{v}_f$ (thick) as functions of r



$\hat{A}_c$ and $\hat{A}_a$ of the two allocation problems increase only very slowly as the discount rate increases.

Looking at Figs. 4.12 and 4.13 we see that in all four models enforcement's share at the steady state increases with r, but that in the unrestricted control problem enforcement in absolute value terms is a decreasing function of the discount rate. To put it in other words, a government which is more farsighted will rely more heavily on treatment measures relative to what is done with respect to enforcement (Fig. 4.13) and vice versa, but if the government is free to decide how much to spend at any time, then the more myopic a government is, the smaller the drug control budget will become .

The dependences of the utility functional values are as one would expect (see Fig. 4.14). First, the more farsighted the government is (which is expressed in low values of r, the higher will be the social costs associated to the drug problem, as what happens at later stages of the drug epidemic is also taken into account. Secondly, if the government is not myopic there will be greater interest in imposing controls on the problem which is the reason

Figure 4.13: $\hat{f}_c$ (thin), $\hat{f}_a$ (regular), and $\hat{f}_f$ (thick) as functions of r

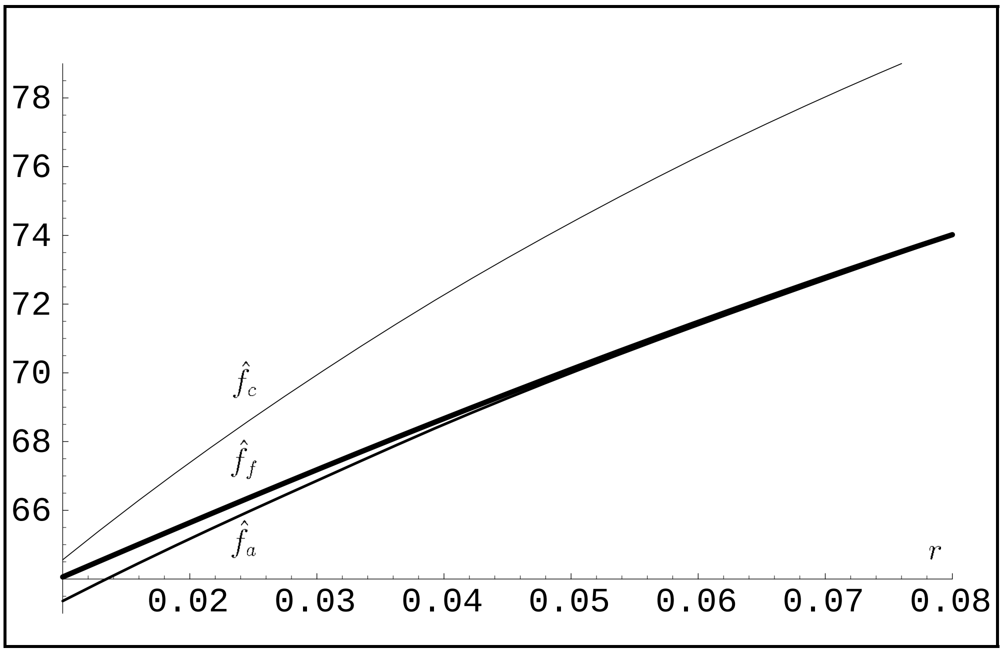
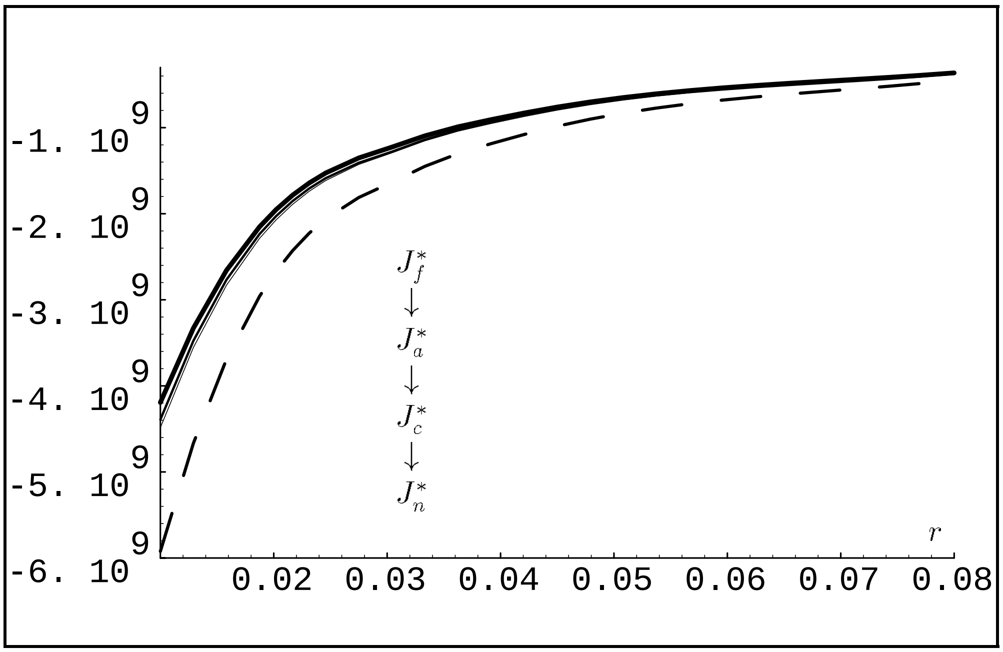


Figure 4.14: J_n^* (dashed), J_c^* (thin), J_a^* (regular), and J_f^* (thick) as functions of r



why especially for farsighted governments it pays to control. For low values of r the utility functional values differ visibly but tend to converge as the discount rate increases.

The magnitude and duration – though not the existence – of the initial spending and resulting spike in prices depend on modeled initiation being independent of prevalence. Comparing the results above with those of Behrens et al. (1997b) suggests that the magnitude and nature of the tendency to rely relatively more on treatment as the epidemic progresses does as well. This suggests analyzing the robustness of the overall policy conclusions not just to the specific parameter values but to the structural assumptions embedded in the model. This can be pursued in at least three ways. First, the results can be compared with those obtained with other models (e.g., Behrens et al., 1997b) that make initiation an explicit function of prevalence. Second, the model's initiation function can be changed to make it depend explicitly on the current number of users; this is being pursued in Chapter 6. Third, we can exogenously impose changes in the initiation rate by changing the parameter k and observe whether such variations overturn any of the inferences drawn so far. This exercise is pursued in the next chapter.

Chapter 5

The Two Phases Problem

5.1 Background and Mathematical Formulation

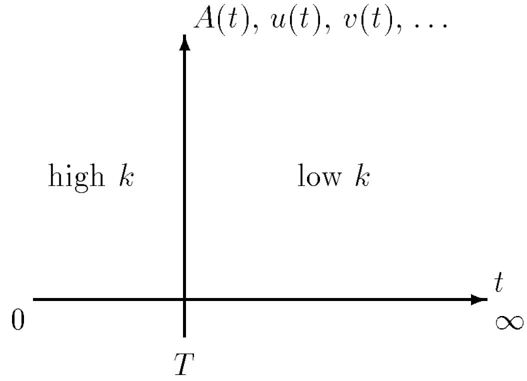
The system dynamics (2.6) of the models described and analysed above (see also Fig. 4.3) is such that – starting with any initial number of users below the steady state value – we have a monotonous increase to that level and stay there forever. However, historically drug epidemics typically die out in the sense that after reaching that climax there follows a decrease to a moderate level of use.

To simulate such an epidemic we could have chosen a function instead of the constant k in the initiation term kp^a which endogenously reproduces this feature. However, this would have caused the models to become far more complex and their analysis to become even harder. Such analyses are deferred to Chapter 6.

Another possibility to reproduce a drug epidemic in our models is to start the dynamics with a low initial number of users A_0 and a high initiation rate expressed by a high value of k (phase 1). Then – after some finite time T – switch to an infinite horizon problem with a low initiation rate k (phase 2). This procedure crudely captures the idea of a drug epidemic and is illustrated in Fig. 5.1.

The main task here is to "glue the two problems together" which can be done by the use of a specific transversality condition. Here, the maximizing condition of the first phase (high $k = k_{high}$, finite time horizon) problem may

Figure 5.1: Illustration of the two phases problem



be written in the form

$$\max_{u,v} - \int_0^T e^{-rt}(\kappa A(t) p\left(A(t), v(t)\right)^{\omega} + u(t) + v(t))dt + e^{-rT} S\left(A(T)\right)$$

(instead of (2.8) with (2.7)) subject to (2.6), where $S\left(A(T)\right)$ is a so-called "salvage value function" which describes the value of being in the state $A(T)$ at the end of the planning horizon. For further descriptions of *problems with free endpoints (terminal states) and a salvage value function* see, e.g., Feichtinger & Hartl (1986), or Leonard & Long (1992). For this problem, at time T we have the transversality condition

$$\lambda(T) = S_A\left(A^*(T)\right) \tag{5.1}$$

where $A^*(T)$ denotes the optimal value of A at time T. A reasonable choice for the salvage value function in our two phases problem is the optimal utility functional value of the second – infinite horizon – problem. Recall that this gives the total social cost, discounted to time T, when the initial state is $A^*(T)$.

The value of the utility functional, however, may be computed as

$$\frac{1}{r} H\left(A(0), u(0), v(0), \lambda(0)\right)$$

(see Feichtinger & Hartl, 1986) so that (5.1) transforms to

$$\lambda^{(k_{high})}(T) = \frac{1}{r} H_A^{(k_{low})}\left(A^*(T), u^{(k_{low})}(0), v^{(k_{low})}(0), \lambda^{(k_{low})}(0)\right)$$

Table 5.1: Additional parameters in the two phases problem; k_{base} denotes the base value of k, $472,618$

Parameter	Value
k_{high}	k_{base}
k_{low}	$\frac{k_{base}}{10}$
T	15

Table 5.2: Steady state values $\hat{A}$ and $\hat{f}$ of the two phases problems as well as optimal utility functional values J^* in absolute values and relative to J_n^* (in percent)

Control	$\hat{A}$	$\hat{f}$	J^*	$\frac{J^*}{J_n^*}$ [%]
n	$1,164,026$	–	$-\ 587,860,242$	100.00
c	$771,781$	0.722	$-\ 472,709,530$	80.41
f	$639,504$	0.687	$-\ 440,536,494$	74.94

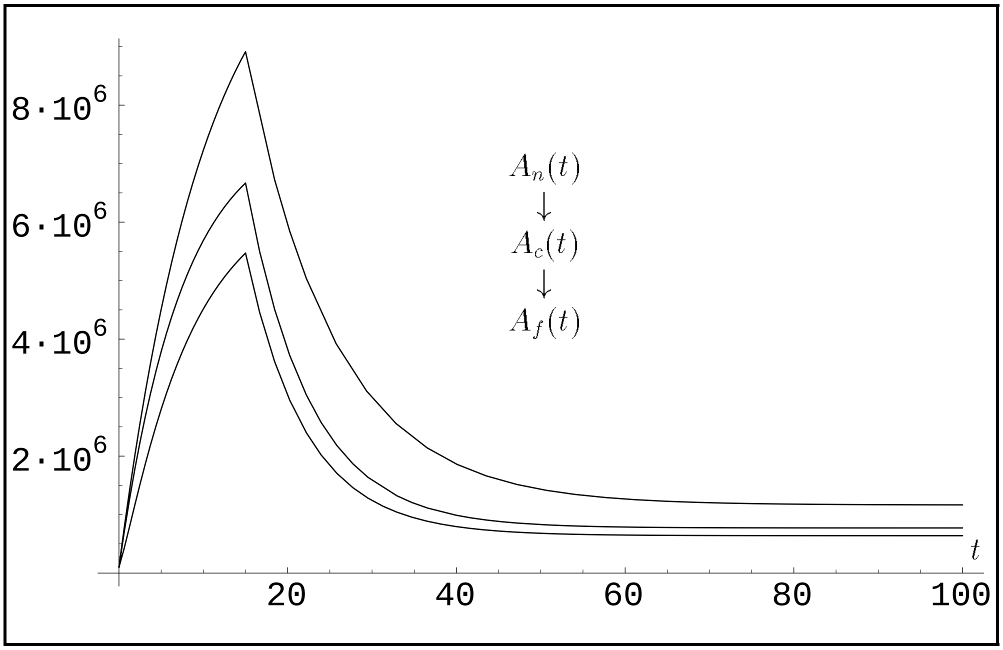
where superscripts (k_{high}) and (k_{low}) denote expressions of the high and low k problems, respectively.

5.2 Analysis

The additional parameters needed for the analysis of the two phases problem where chosen as given in Table 5.1. For technical reasons, the optimally controlled allocation problem has been excluded from the analysis described below.

The steady state values $\hat{A}$ and $\hat{f}$ of the two phases problems as well as the optimal utility functional values J^* in absolute values and relative to the base value of no control are given in Table 5.2. Comparing these percent values with those in Fig. 4.2 we see that the optimal utility functional value in the constant fraction control problem relative to the uncontrolled case remains nearly the same in the two phases problem, whereas there is a 1.6 % improvement in the value of the unrestricted optimal control model (which

Figure 5.2: Time paths $A_n(t)$ (top), $A_c(t)$ (middle), and $A_f(t)$ (bottom) in the two phases problem for $t \in [0, 100]$



is still not very much).

Fig. 5.2 gives the time paths of the numbers of users in the three problems to be considered. Again, it holds that $A_n(t) > A_c(t) > A_f(t)$ $(\forall t)$. Except for the kinks at time $T = 15$, which are caused by the sudden change from k_{high} to k_{low}, the graphs look like those produced by the one phase problem.

The temporal evolutions both of treatment and enforcement spending are illustrated in Figs. 5.3 and 5.4, respectively. In the constant fraction control problem (thick curves), u and v as functions of time t are continuous, because the optimal fraction f which determines the budgets for enforcement and treatment spending – fGA and $(1-f)GA$, respectively – is determined once and for all, and A is continuous. This is not true in the unrestricted control problem. For that problem, u behaves rather "well", but in the time path of v we see a dramatic drop at time T. Nevertheless, there are some similarities in the time paths of the constant fraction and the unrestricted control problems.

Figure 5.3: Time paths $u_c(t)$ (thick) and $u_f(t)$ (thin) in the two phases problem for $t \in [0, 100]$

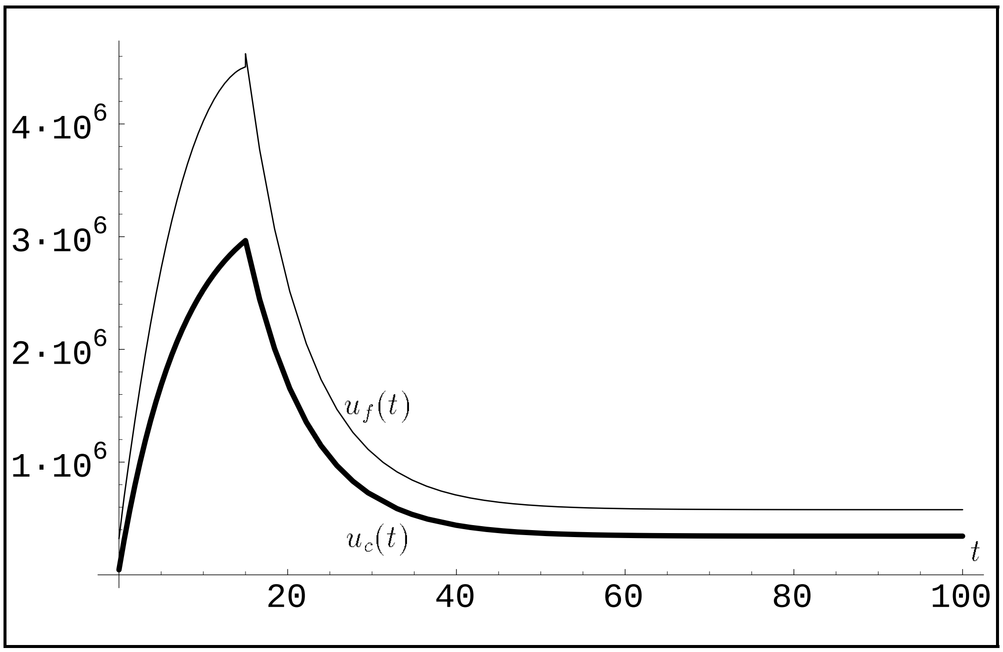


Figure 5.4: Time paths $v_c(t)$ (thick) and $v_f(t)$ (thin) in the two phases problem for $t \in [0, 100]$

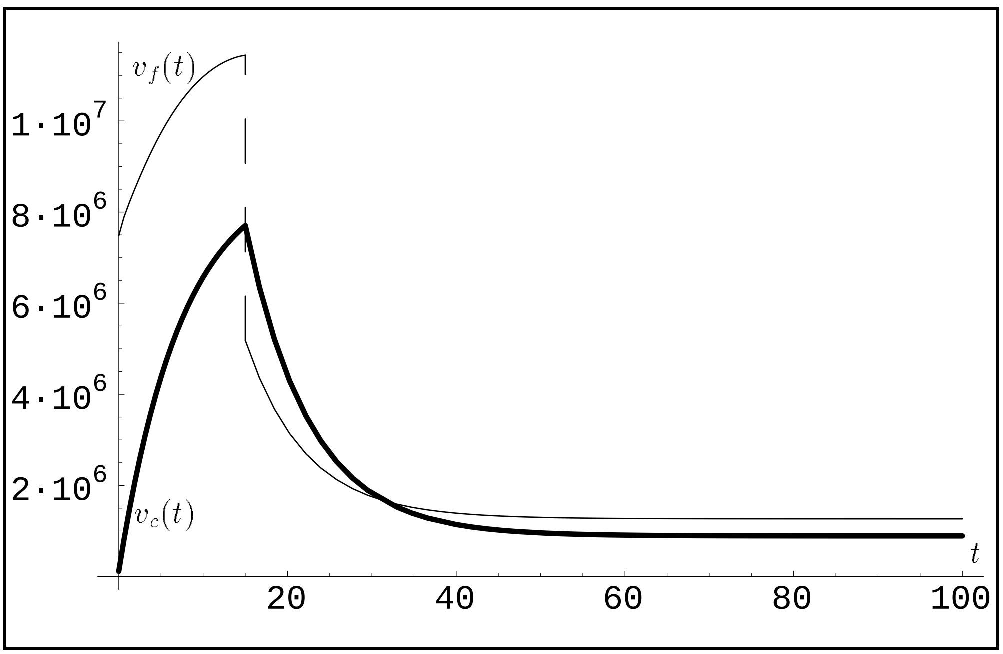
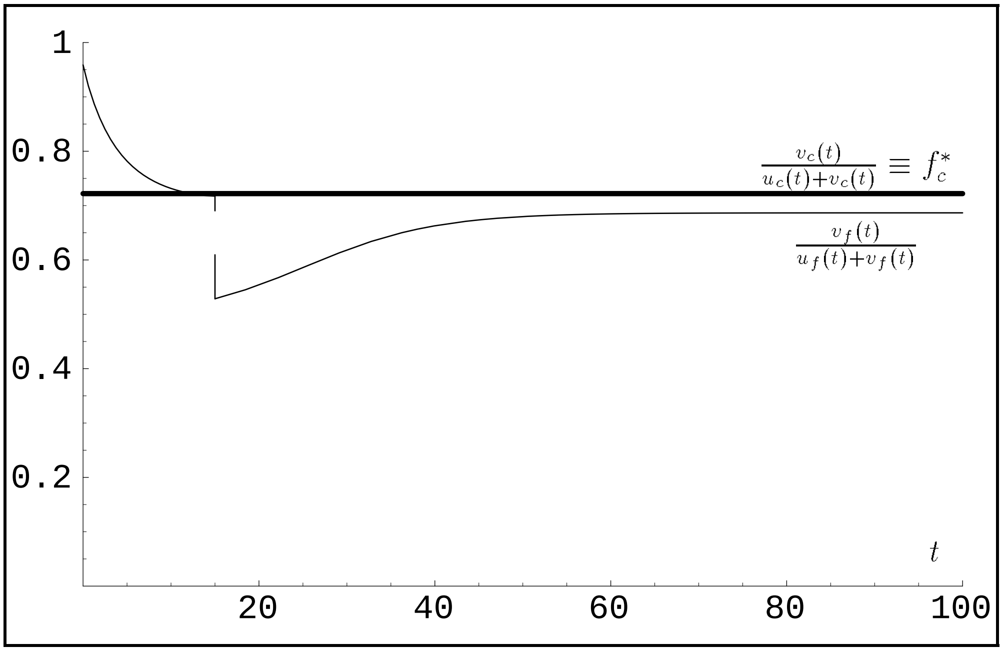


Figure 5.5: Time paths $\frac{v_c(t)}{u_c(t)+v_c(t)} \equiv f_c^*$ (thick) and $\frac{v_f(t)}{u_f(t)+v_f(t)}$ (thin) in the two phases problem for $t \in [0, 100]$



Enforcement's share along the optimal path is given in Fig. 5.5. In the constant fraction control problem, some "intermediate" value is chosen for the optimal shares between treatment and enforcement.

In this chapter we solved the model with the initiation parameter k which is at a constant, high level for a certain number of years and then decreases to some lower value and stays at that lower value forever. This is very crude way of simulating the historical trend in initiation. The results were entirely consistent with what one would expect given the analysis in Chapters 1–4. The absence of any surprises is reassuring. It implies that the qualitative conclusions drawn earlier are robust with respect to even abrupt, discontinuous, exogeneously imposed changes in initiation. The only difference we observed is that, quantitatively, it seems that the value of pursuing an optimal dynamic control relative to some simpler heuristic is greater when initiation changes than when it is constant, which makes intuitive sense. Flexibility is more important when the environment is changing than when it is stable.

Chapter 6

Modeling the Feedback Effect of Prevalence on Initiation

6.1 The Idea

Up to now we have assumed that the initiation term in the system dynamics, $kp(A,v)^a$, depends on prevalence (the number of users) only indirectly through price. However, most people who start using drugs do so through contact with a friend or sibling who is already using. In other words, one can say that current users tend to "recruit" new users, so that talking about a "drug epidemic" makes sense. On the other hand, the question arises why initiation does not increase monotonically. One explanation for that is given by Musto (1987) who argued that, in addition to that recruitment effect, people become aware also of the adverse effects of drug use which acts as a deterrent.

It would be easier to model these ideas in a two state model, say, where we distinguish between "light" and "heavy" users (cf. Everingham & Rydell, 1994). In such a model, the light users are those who are capable of recruiting new users, whereas the heavy users with all their visible problems suppress initiation into drug use (see, e.g., Behrens et al., 1997a,b).

If we want to incorporate these feedback effects into our one state model, we have to let initiation depend on prevalence (i.e., the number of users, A) in such a way that both the positive and the negative influences are included by using only this one state A. One way to do that is to use an

Table 6.1: Additional base parameter values in the problems with state dependent initiation term

Parameter	Value
α	0.3
k	4272.0469

additional initiation function factor depending on A, $I = I(A)$, which has positive influence on initiation over the whole course of the epidemic, but with marginally decreasing intensity to take into account the negative influence by heavy users which becomes the more visible the more users there are. The most obvious and probably simplest function with these properties is a power function of A,

$$I(A) = A^{\alpha} \quad \text{with} \quad 0 < \alpha < 1.$$

Using that, the state dynamics equation (2.6) transforms to

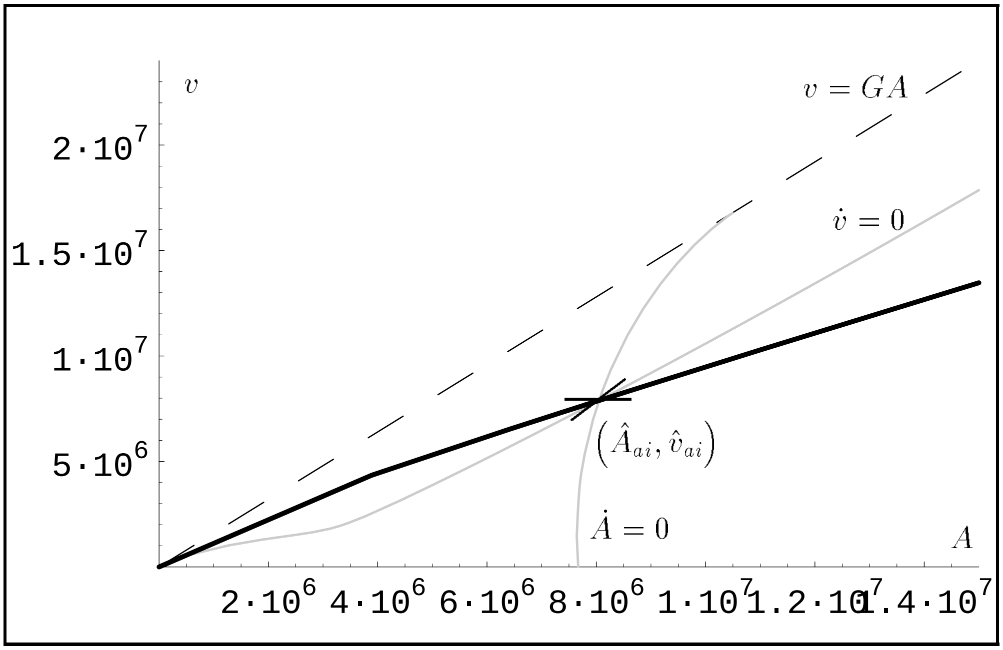
$$\dot{A} = kp^a A^{\alpha} - c\beta^z A - \mu p^b A. \tag{6.1}$$

6.2 Analysis

The analysis is analogous to that of Chapter 3. Properly speaking, the original system dynamics (2.6) is just a special case of the new state equation (6.1) if we set $\alpha = 0$. The two-dimensional system to be analysed consists of the state equation (6.1) and the differential equation for the control v as given in (3.10) for the optimal control allocation problem and (3.13) for the unrestricted optimal control problem, respectively, but – due to the change in the initiation term – with new expressions for $N(\lambda)$, $Z(\lambda)$, $H_A^{(1)}$, $H_A^{(2)}$, $Z(\lambda_A)$, $Z(\lambda_v)$, and $\dot{A}$ (use Appendices A.4 and A.5 with $I(A) = A^{\alpha}$).

For what follows we have to specify the value of the new parameter α and recalculate the value of the initiation proportionality constant k (see Table 6.1) so that the whole initiation term in the base year still yields $1,000,000$ new users. We choose $\alpha = 0.3$ because in the optimal control allocation problem that results in a steady state value of A which is in the range of what we have in the original allocation problem with initiation term kp^a.

Figure 6.1: Phase portrait in the A-v-plane for the optimal control allocation problem with the state dependent initiation term $kp^a A^\alpha$; again, the dashed line represents the budget constraint border line $v = GA$, the grey curves represent the $\dot{A} = 0$ and $\dot{v} = 0$ isoclines, and the thick black curves are the stable manifolds of the saddle point equilibrium $\left(\hat{A}_{ai}, \hat{v}_{ai}\right)$



To draw a distinction between the values calculated in this chapter and those calculated in the previous ones, with indices $_{ai}$ and $_{fi}$ we refer to values of the optimal control **a**llocation and unrestricted (**f**ree) optimal control problems, respectively, with state dependent **i**nitiation term $kp^a A^\alpha$.

6.2.1 The Optimal Control Allocation Problem

Looking at Fig. 6.1, which is a phase portrait in the A-v-plane, we see that including the factor A^α in the initiation term does not change at all the qualitative behaviour of the controlled system dynamics. Also, the steady state values $\hat{A}_{ai}$, $\hat{u}_{ai}$, $\hat{v}_{ai}$, and $\hat{f}_{ai}$ as well as the optimal utility functional value J^*_{ai} given in Table 6.2 are comparable to those of the original problem

Table 6.2: Steady state values $\hat{A}_{ai}$, $\hat{u}_{ai}$, $\hat{v}_{ai}$, $\hat{f}_{ai}$, and the optimal utility functional value J^*_{ai} for the optimal control allocation problem with state dependent initiation term $kp^a A^\alpha$

$\hat{A}_{ai}$	$\hat{u}_{ai}$	$\hat{v}_{ai}$	$\hat{f}_{ai}$	J^*_{ai}
8,049,026	4,990,502	7,887,939	0.612	- 837,908,295

(cf. Table 3.5).

Not surprisingly, the qualitative shapes of the time paths $A_{ai}(t)$, $u_{ai}(t)$, $v_{ai}(t)$, $f_{ai}(t)$, and $B_{ai}(t)$ are not very different than those in the problem without a state dependent initiation term. That is why we dispense with showing them here. However, for reasons of completeness, a generalized phase diagram with the optimal control values both of v and of u is given in Fig. 6.2.

6.2.2 The Unrestricted Optimal Control Problem

In contrast to the allocation problem, the results of the analysis of the unrestricted optimal control problem turn out to be completely different from those of the original problem, both qualitatively and quantitatively. This can be seen in Fig. 6.3 which is a phase portrait in the A-v-plane. As we can see, there not only exists a ("**h**igh") saddle point equilibrium, $\left(\hat{A}^{(h)}_{fi}, \hat{v}^{(h)}_{fi}\right)$, as in all the cases above, but also a second ("**l**ow") equilibrium, $\left(\hat{A}^{(l)}_{fi}, \hat{v}^{(l)}_{fi}\right)$, which is an unstable focus. This implies a different optimal policy which is described in what follows.

The first idea here is that again the stable manifolds of the saddle point equilibrium should be the optimal paths. This is true for "many" initial states A_0. However, along the stable manifold left of the steady state value $\hat{A}^{(h)}_{fi}$ there is a minimum level of users, A_{min}, which is greater than zero, so it is not possible to jump on to that trajectory for initial states A_0 which are smaller than A_{min}.

In other words, it is not clear at first sight which path is optimal left of A_{min}. This question can be answered by introducing a lower limit of users, $\underline{A}$, which represents the idea that – no matter how effective the law enforcement

Figure 6.2: Generalized phase portrait with the stable manifolds of the diagrams both in the A-v-plane (black) and the A-u-plane (grey) for the optimal control allocation problem with the state dependent initiation term $kp^a A^\alpha$; the vertical black line indicates the steady state value $\hat{A}_{ai}$

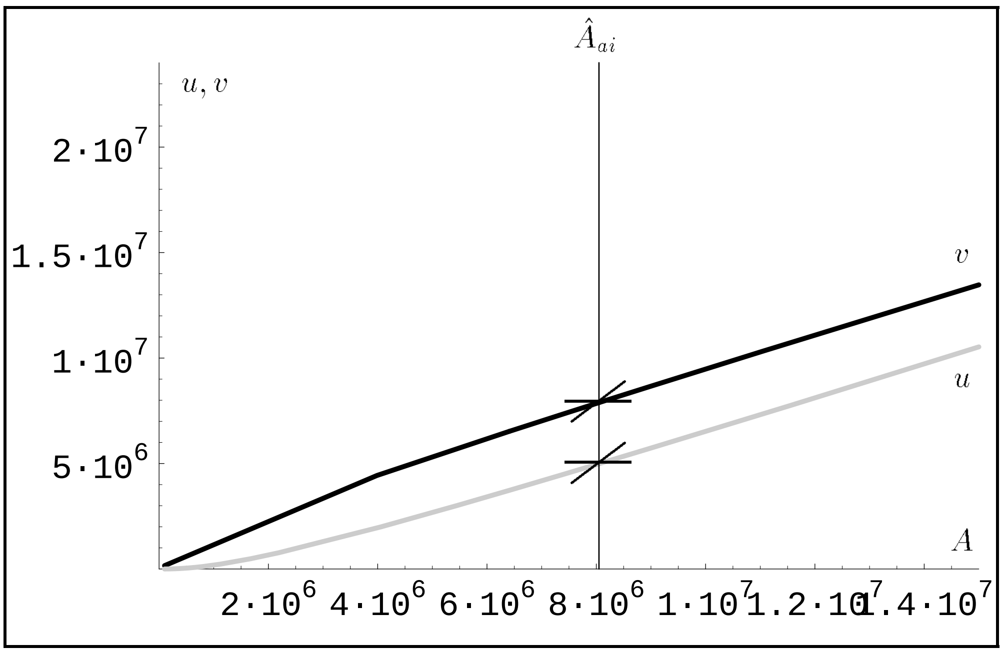
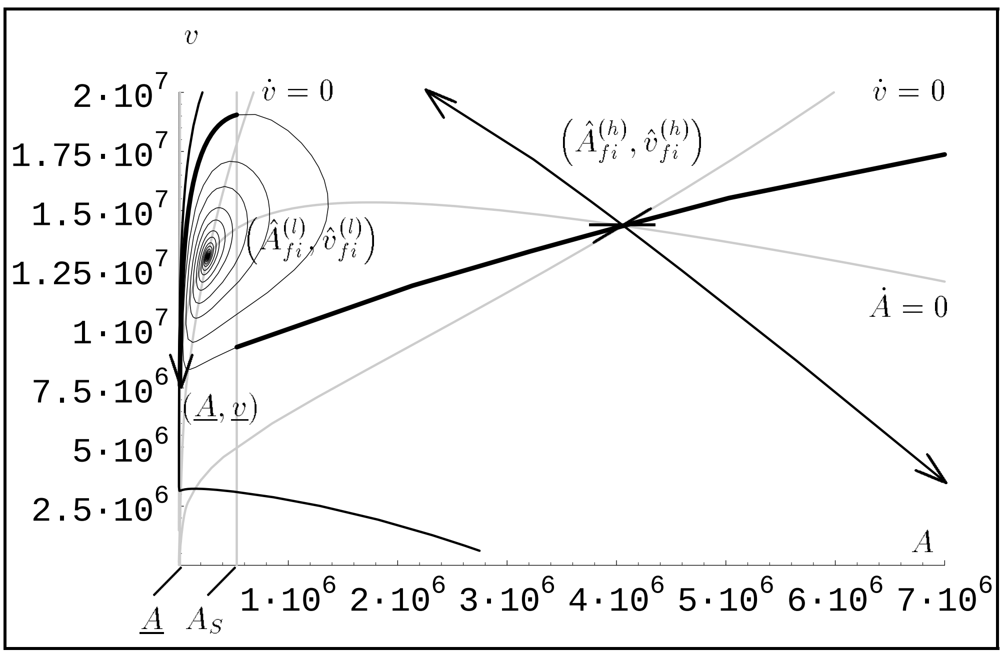


Figure 6.3: Phase portrait in the A-v-plane for the unrestricted optimal control problem with the state dependent initiation term $kp^a A^\alpha$; the two grey vertical lines indicate the minimum level $\underline{A} = 10,000$ and the Skiba threshold $A_S = 529,117$, respectively; there are two $\dot{v} = 0$ and one $\dot{A} = 0$ isoclines (grey curves); the optimal trajectories are given by the thick black curves: for initial states below the Skiba point the movement is towards the minimum steady state $(\underline{A}, \underline{v})$, above A_S the optimal trajectories lead to the "high" equilibrium $\left(\hat{A}^{(h)}_{fi}, \hat{v}^{(h)}_{fi}\right)$; the low steady state $\left(\hat{A}^{(l)}_{fi}, \hat{v}^{(l)}_{fi}\right)$ is unstable



agency is – there will always be a small but positive number of users. With that "level of undetectable users" one can show that the point $(\underline{A}, \underline{v})$ becomes another equilibrium value, where $\underline{v}$ is given by the intersection of the vertical line $A = \underline{A}$ and the isocline $\dot{A} = 0$. This steady state is approached along that trajectory which spirals out of $\left(\hat{A}_{fi}^{(l)}, \hat{v}_{fi}^{(l)}\right)$ and leads to $(\underline{A}, \underline{v})$.

Summing up we can say that there are three equilibria, but only two of them are stable in an optimal control theoretic sense, i.e., the "minimum" steady state $(\underline{A}, \underline{v})$ and the high steady state $\left(\hat{A}_{fi}^{(h)}, \hat{v}_{fi}^{(h)}\right)$. If we look at the two trajectories which lead to these equilibria it is obvious that for very low levels of A_0 one can only jump on to the path leading to the minimum level, whereas for rather high initial states it is only possible to jump on to the left stable manifold ending in the high equilibrium. However, there is also a broad range of initial states A_0 where it is not clear in the first instance which of the two trajectories is optimal, i.e., which path leads to a higher utility functional value.

Fig. 6.4 gives the optimal utility functional values along the two trajectories under consideration as functions of the initial state A_0. We see that there is a unique point, $A_S = 529,117$, so that left of A_S the trajectory leading to the minimum equilibrium $(\underline{A}, \underline{v})$ yields higher utility functional values, whereas for all initial states which are greater than A_S the left stable manifold of the high equilibrium $\left(\hat{A}_{fi}^{(h)}, \hat{v}_{fi}^{(h)}\right)$ leads to lower total social costs, i.e., depending on the initial state A_0 it is either optimal to "eradicate" drug use or to approach a high equilibrium level of use. The threshold A_S defining the two basins of attraction is a so-called Skiba point (Skiba, 1978; cf. also Feichtinger et al., 1997).

Table 6.3 gives the minimum and high steady state values as well as the optimal utility functional value for the unrestricted optimal control problem with a state dependent initiation term.

Three things are interesting with these values. First, comparing the values of the two – high and minimum, respectively – equilibria, we see that treatment and enforcement have opposite roles in either steady state, i.e., treatment is funded more generously at $\underline{A}$, whereas enforcement has a greater share at $\hat{A}_{fi}^{(h)}$. In Fig. 6.5 we have the optimal values of treatment and enforcement along the whole optimal paths. We see that

$$u \left\{ \begin{array}{c} > \\ < \end{array} \right\} v \text{ if } A \left\{ \begin{array}{c} < \\ > \end{array} \right\} A_S,$$

Figure 6.4: Optimal utility functional values as functions of the initial state A_0 along the trajectory leading to the minimum equilibrium $(\underline{A}, \underline{v})$ (grey, "$\underline{J}^*$") and along the left stable manifold of the high equilibrium $\left(\hat{A}_{fi}^{(h)}, \hat{v}_{fi}^{(h)}\right)$ (black, "$J^{(h)*}$") with $A_0 \in [175,000; 1,300,000]$; for initial states A_0 below A_S, $\underline{J}^*$ exceeds $J^{(h)*}$ and vice versa

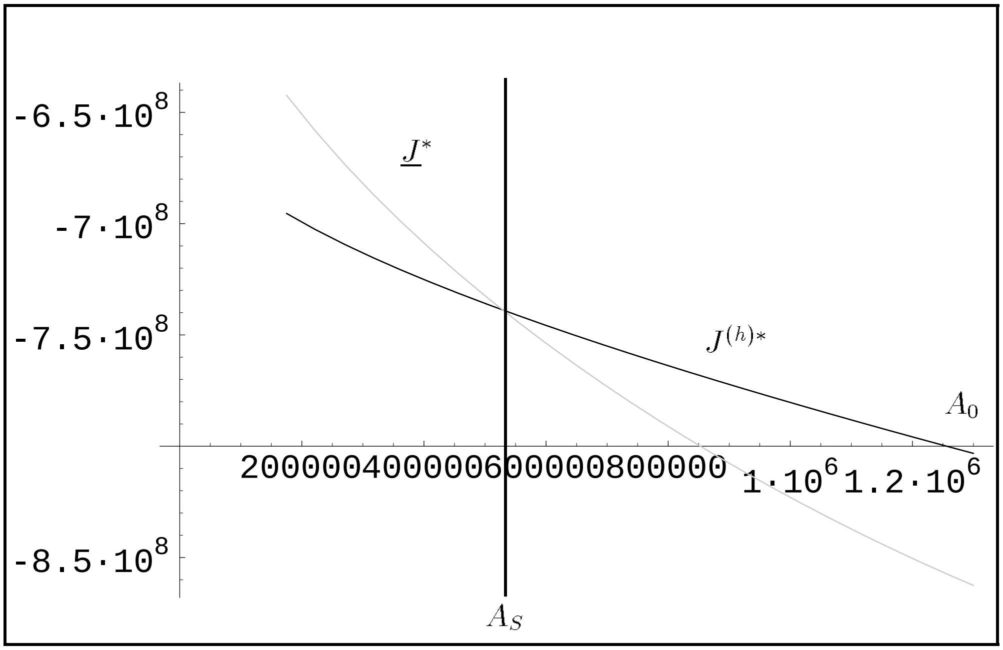
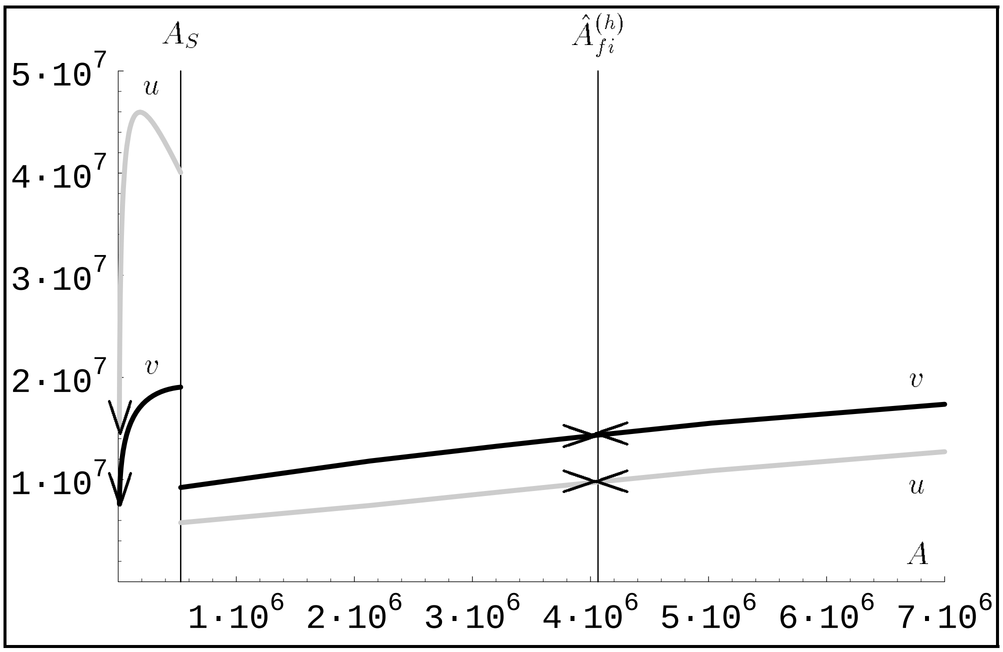


Table 6.3: Steady state values and the optimal utility functional value for the unrestricted optimal control problem with state dependent initiation term kp^aA^α

$\underline{A}$ $\hat{A}_{fi}^{(h)}$	$\underline{u}$ $\hat{u}_{fi}^{(h)}$	$\underline{v}$ $\hat{v}_{fi}^{(h)}$	$\underline{f}$ $\hat{f}_{fi}^{(h)}$	J^*
$10,000$	$14,867,020$	$7,601,528$	0.338	- 612,671,721
$4,063,819$	$9,765,640$	$14,356,905$	0.595	

Figure 6.5: Treatment (grey) and enforcement (black) as functions of A along the optimal paths; the left and right vertical lines indicate the Skiba threshold A_S and the high steady state value $\hat{A}_{fi}^{(h)}$, respectively



i.e., this property is not restricted to the values at the steady states. To put this in other words, treatment turns out to be more cost-effective when the government is outflow-oriented, while enforcement measures are more cost-effective along the consumption-oriented path leading to the high equilibrium.

The second remarkable result that follows from Table 6.3 and – in more depth – from Fig. 6.5 is that treatment and enforcement are on a very high level left of the Skiba point where one approaches the minimum level of users, $\underline{A}$. In other words, the total budget that has to be spent if one wants to end up with a low number of users has to be extremely high per user. Even in absolute terms, one spends about as much on controls after the problem has been "eradicated" as one would if the epidemic were allowed to grow to its high equilibrium, and in the transition period when one is driving the epidemic toward "eradication", one spends much more. This might be the reason why for the allocation problem of the preceding subsection we did not have the "choice" of eradicating drug use. The budget constraint $B = GA$ did not allow spending so much, especially when there were few users. This comparison shows that how a government handles the budget question can matter a lot.

That the budget question matters a lot not only in terms of how many users one ends up with but also with respect to the total budget spent is the third thing that is interesting when looking at Table 6.3. Comparing the optimal utility functional value J^* there with that of the allocation problem in Table 6.2, we see that here the government can save a lot (26.88 %) by not binding the budget to be proportional to the number of users. Note that the J^*'s where again computed under the assumption that the government starts to control at $A_0 = 100,000$.

One thing that has to be discussed in more detail is the minimum level of users, $\underline{A}$. The government has some, but not complete control over $\underline{A}$, i.e., the government can influence the minimum level of users which is optimal for initial states less than the Skiba threshold. Although it probably does not matter in a big country like the United States if the steady state level of drug users is, say, $5,000$, $10,000$, or $20,000$, there may be a difference in the total social costs computed over the whole planning horizon.

Fig. 6.6 is a plot of the optimal total spending, $u + v$, as a function of A for initial states which are smaller than the Skiba threshold (i.e., $A_0 < A_S$) with $\underline{A}$ being defined by $5,000$ (thin), $10,000$ (regular), and $20,000$ (thick),

Figure 6.6: The optimal total spending $(u+v)$ as a function of A for initial states less than the Skiba threshold with $\underline{A} = 5,000$ (thin), $10,000$ (regular), and $20,000$ (thick), respectively; the curves range from $\underline{A}$ to A_S

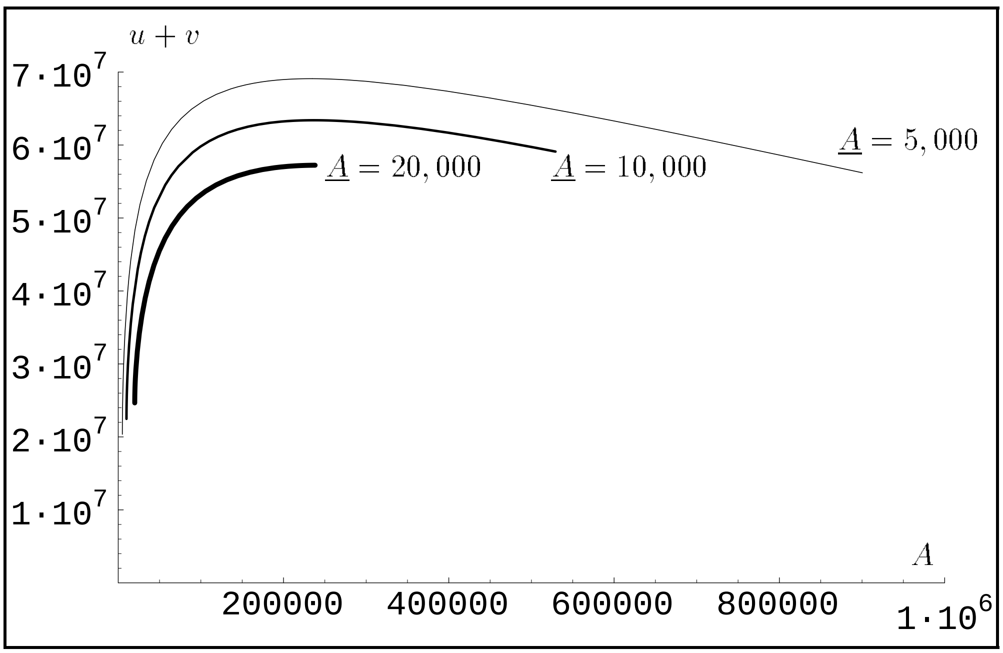


Table 6.4: Values of A_S and J^* as well as time τ which is needed to reach $\underline{A}$ starting at $A_0 = 100,000$ for different values of $\underline{A} = 5,000$, $10,000$, and $20,000$

$\underline{A}$	A_S	J^*	τ
$5,000$	$900,482$	$-567,972,999$	2.195
$10,000$	$529,117$	$-612,671,721$	2.396
$20,000$	$238,274$	$-656,707,935$	2.579

respectively. We see that along these trajectories the total budget spent per user is greater, if $\underline{A}$ is chosen to be smaller, which makes intuitive sense. However, this is only one part of the story. Looking at Table 6.4 we see that it takes longer to reach the steady state if $\underline{A}$ is bigger. In addition, the larger $\underline{A}$ is, the more users consume in steady state and the more costly it is to maintain that steady state, presumably due to the fact that initiation increases when the number of users increases. Also, the smaller $\underline{A}$ is, the larger A_S is. That is, the smaller the minimum number of drug users to which enforcement and treatment can drive the system, the larger the critical number of users is which determines whether it is better to use very aggressive controls to all but eradicate use or whether it is better to use more limited controls that ameliorate at the margin the expansion of drug use toward its high volume equilibrium. Summing up, we see that for higher values of $\underline{A}$ the longer transition from $A_0 = 100,000$ to $\underline{A}$ and the higher steady state values both of A and $u + v$ outweigh the lower budget which is spent during the transition from $t = 0$ until $t = \tau$ (see the column of the J^*'s in Table 6.4). Another interesting feature that follows from the values of A_S in Table 6.4 is that the Skiba threshold is greater if a smaller value of $\underline{A}$ is chosen, which means that the government has more time to start the controls. Hence, with the parameters used in this analysis, there is no reason for the government to prefer a "high" minimum level of users $\underline{A}$. Note, however, that these results might change if, e.g., the weights κ and G in the utility functional are chosen in a different way.

Chapter 7

Discussion

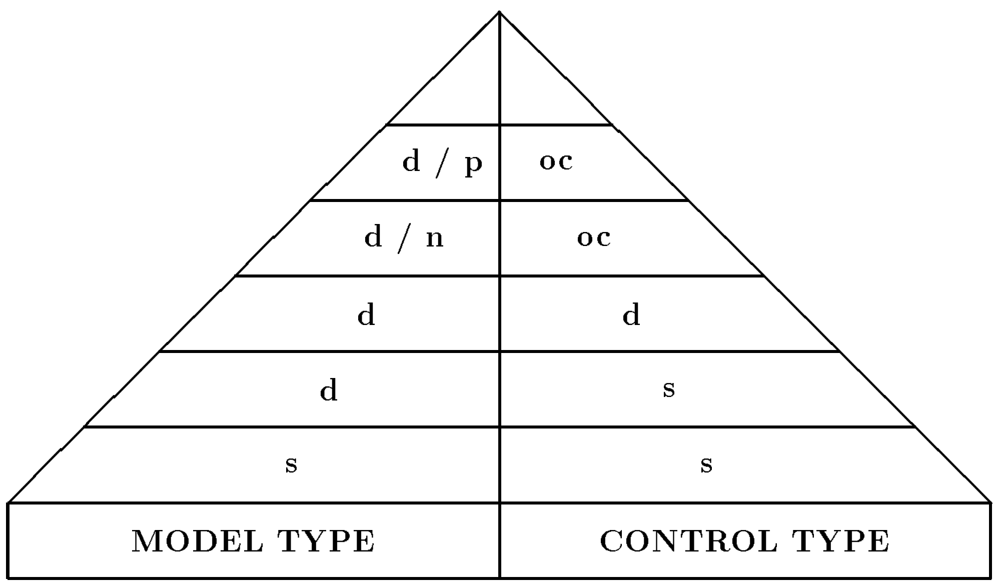
This paper introduced a dynamic (optimal control) model of illicit drug consumption where the government seeks to minimize the total social costs consisting of the social costs caused by drug consumption and the expenditures for two control mechanisms, i.e. treatment and price-raising enforcement.

At first sight, the model presented here seems to be rather simple. However, taking a look at the broad literature on controlling illicit drugs, one can see that with this work we are still at the very beginning of a new research period. Most models on illicit drug control are either purely *static* (see, e.g., Caulkins, 1993) or *dynamic* but with *static* controls (Baveja et al., 1993). There are a few *dynamic* models that include *dynamic* controls (Baveja et al., 1993), but only a few of them are *optimal control* models like the one presented in this paper (Kort et al., 1996). Assuming this (without any doubt) incomplete classification of mathematical models on illicit drug control – illustrated in Fig. 7.1 – our model could be ranked one level above that of Kort et al. (1996), because in addition to using an optimal control approach, for the analysis of Chapters 3–6 we also chose the parameter values according to (real world) data from the U.S. cocaine epidemic, which – to the best of our knowledge – has not been done before (apart from that, we implemented *two* control mechanisms – treatment and enforcement).

We first review some of the important insights from the analysis of our model, and then conclude this paper with directions for future research.

It turns out that it is of enormous importance when a government starts to impose controls on a drug problem. In particular, the later the controls start, the higher will be the discounted stream of total social costs. One can

Figure 7.1: (By no means complete) classification of mathematical models on illicit drug control; abbreviations are: **d**ynamic, **n**ot parameterized, **o**ptimal **c**ontrol, **p**arameterized, **s**tatic; complexity increases from bottom to top



compute the cost per year of delaying initiation of controls and from that estimate how much one should be willing to spend on a monitoring system that would let the government detect a new drug epidemic earlier, which allows controls at early stages resulting in fewer total social costs.

Further, the question of how much the government should spend on drug control depends strongly on how high the social costs caused by drug consumption are assumed to be, which makes intuitive sense. For any estimate of the per gram costs, there exists one value of the per user budget that would be optimal. If the per gram costs are lower, the government would be better off cutting the size of the drug control budget, whereas for high social costs associated with drug consumption, it would pay to increase the budget. To put this in other words, the question of whether a society should favor a more or less aggressive drug control program will depend on the – probably subjective – estimation of the costs that come with the consumption of illicit drugs.

We also saw that – in addition to a high level of users – there can be a second, low steady state value, which is approached in the long run. In that case there exists a so-called Skiba threshold so that for initial states above the Skiba point the optimal trade-off between social costs implies an equilibrium with a high level of users, while below the threshold the optimal controls should drive the system to a negligible number of users. It is important to note, however, that for the existence of a low steady state, it is necessary that the expenditures for treatment and enforcement not be restricted to be proportional to the size of the epidemic; it is shown that if the budget is proportional to the number of users, the optimal policy cannot "eradicate" drug use. This makes life hard for politicians: although with high levels of treatment and enforcement it can be possible to approach a low level of use which also implies lower total social costs over the whole planning horizon, it will be difficult to justify those high expenditures, which are necessary for that purpose, if the problem is hardly visible.

We conclude this paper by calling the reader's attention to a series of extensions of the models presented here that should be taken into consideration in order to go a step further than does the present work. These extensions include:

- **Other allocation problems.** In the allocation problems presented above we assumed that the budget at any time is proportional to the number of users at that time. One can imagine that the budget could also be constant at least over some time period. In mathematical terms, the budget constraint (2.11) simplifies to

$$u(t) + v(t) = G \quad \forall t,$$

and the budget term in the utility functional disappears.

In a more sophisticated approach, one would model the budget B as a second state variable:

$$\dot{B}(t) = -u(t) - v(t) + r_2 B(t) + \varphi\left(A(t)\right),$$

where r_2 gives the interest rate, and $\varphi(A)$ represents replenishment of the budget which depends on the size of the problem characterized by the other state variable, A. In this case, one should assume

$$u(t) + v(t) \leq B(t) \quad \forall t.$$

Another variant of the allocation problem is sketched below ("delays").

- **Extensions of the initiation term.** The base model assumed a constant initiation rate, k, which is reduced by enforcement-induced price increases. This is a rather naive assumption, and we saw that it does not allow for a decline in the number of users. We have already introduced the idea of a drug epidemic by investigating the two phases problem where in the first (finite time) phase we have high initiation expressed by a high value of k, whereas the second phase is characterized by a lower initiation rate (i.e., lower k). Apart from that, in the preceding chapter we analysed our model with an initiation term, which depends directly on the number of users (i.e. not only through price). Still there are several other initiation terms that would be worth being analysed like, e.g., a logistic initiation:

  $$kp^a \mapsto kA\left(\bar{A} - A\right)p^a$$

  with a suitable constant $\bar{A}$.

- **Stochasticity.** We even could assume the initiation constant becomes a stochastic variable which reflects the idea that initiation changes (stochastically) with the spirit of the age: in 1968, e.g., people were much more susceptible to drugs than in the early nineties, say.

- **Delays.** In this paper we assumed that any control influences the system without any delay. It might be more realistic to let the price change some time *after* enforcement has been in-/decreased. Analogously, treatment in reality does not cause a drug user to stop using drugs immediately.

  As already mentioned above, an interesting variant of the allocation problems could be modeled with delaying the time when a certain budget will be at disposal. In that case, the budget constraint (2.11) becomes
$$u(t) + v(t) = GA\,(t - \tau)$$
  with $\tau \geq 0$ denoting the delay.

  However, including delays in optimal control models is most challenging from the analyst's point of view.

- **Prevention.** Although currently in the U.S. prevention's share in the drug control budget is relatively low compared to enforcement expenditures, introducing a third control should be most interesting. In a straightforward approach prevention would be included in the initiation term to yield
$$kp^a \mapsto kp^a \Psi\left(\frac{w}{A}\right),$$
  where w denotes prevention expenditures, $\Psi(0) = 1$, $\frac{\partial \Psi}{\partial w} < 0$, $\frac{\partial^2 \Psi}{\partial w^2} > 0$, and $0 < \lim_{w \to \infty} \Psi < 1$.

- **Heterogeneous aspects.** In this paper we assumed that the population of drug users is homogeneous in the sense that the consumption rate is equal for all users. This, of course, can only be interpreted as a crude approximation of real life. Ideally one would model the whole spectrum of consumption behaviour, from occasional use in small amounts up to frequent use in large amounts, but data limitations make that infeasible, apart from the fact that too many levels of consumption would make the model extremely hard to analyse. Recognizing this tension, Everingham and Rydell (1994) suggest that, at least for cocaine, a simple dichotomous distinction between "light" and "heavy" users is sufficient (see Behrens et al., 1997a,b, for a first approach in that direction).

- **Networks.** In contrast to the models presented here, in which we assumed that all people are aggregated in one group (i.e., number of users, A), it would be interesting to describe the black market situation of drug selling by using a social network model. This would allow one to explicitly model association between people, knowledge that other people are involved with drugs, and, e.g., who supplies whom with drugs in that network.

Appendix A

Technical Details

A.1 Derivation of the Base Parameter Values

The base parameter values were chosen under the assumptions that monetary units are in \$ 1,000 and that the base year is 1992, and in consideration of Rice et al. (1990), Everingham & Rydell (1994), Rydell & Everingham (1994), Rhodes et al. (1995), Johnson et al. (1996), Miller et al. (1996), the report by the Office of National Drug Control Strategy (1996), Rydell et al. (1996), Caulkins et al. (1997), and ONDCP (1997). The following (additional) assumptions were made:

- Rhodes et al. (1995) estimate that there were 6,680,000 cocaine users in the U.S. in 1992 and 6,292,000 in 1993. We take $A = 6,500,000$ as an average of these figures.

- Rydell & Everingham (1994) report total cocaine control spending of about \$ 13.0 billion (in 1992 dollars), with domestic enforcement (v) receiving \$ 9.5 billion, interdiction receiving \$ 1.7 billion and \$ 0.9 billion each going to source country control and treatment (u). Our model focuses on domestic enforcement and treatment, so we take total spending per user to be

$$G = \frac{\$\ 9.5 \text{ billion} + \$\ 0.9 \text{ billion}}{6,500,000} = \$\ 1,600 \text{ per user.}$$

- ONDCP (1997) reports a price per pure gram for purchases of 5 oz. or less of \$ 106.73 in 1992.

- Rydell & Everingham (1994) estimate 1992 consumption at 291 metric tons.

- Rydell & Everingham (1994, p. 38) report 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, implying an average cost per gram of \$ 67.6 (in 1992 dollars). The Rice numbers are conservative in their estimation of the costs of crime, particularly in light of Miller et al. (1996), so we take \$ 100 per gram as the base case number.

- The average annual rate of initiation reported by Johnson et al. (1996) between 1972 and 1992 was $1,034,571$ and between 1982 and 1992 it was $1,061,091$, so we choose k to yield $1,000,000$ initiations per year as reflecting typical initiation rates.

- Rhodes et al. (1995) report that there were $4,331,000$ light cocaine users and $2,349,000$ heavy users in 1992. According to Everingham & Rydell's (1994) model, about 15 % of the former and 2 % of the latter cease use each year without the benefit of treatment. That implies that in base case conditions, the outflow rate is about

$$4,331,000\frac{15}{100} + 2,349,000\frac{2}{100} = 696,630$$

  or about $700,000$ users per year.

- From Rydell et al. (1996) we get that around 30 % of the heavy cocaine users receive treatment in any given year. The Office of National Drug Control Strategy (1996) states that around one-third of all users are heavy users, and Rydell et al. (1996) estimate that 13.2 % of heavy users treated leave heavy use because of that treatment. Those figures imply that

$$\frac{30}{100}\frac{6,500,000}{3}\frac{13.2}{100} = 85,800$$

  users stop drug use each year due to treatment.

- The values for a, b, and ω together reflect a belief in a long term elasticity of demand of -1 (base case in Caulkins et al., 1997), that the short term elasticity is half the long-term elasticity (Rydell & Everingham, 1994), and that the long term elasticity portion can be divided equally between effects on initiation and exit (Rydell & Everingham, 1994). I.e., $a = -0.25$, $b = 0.25$, and $\omega = -0.5$.

- $\delta = \epsilon = 0.001$ are just small constants inserted to avoid zero dominators; they have no material effect on the results.

- The exponent z associated with the efficiency of treatment function, $\beta(A, u)$, should be between 0 and 1 to take into account the fact that efficiency increases with u, but every extra dollar spent has less impact on the outflow of users than the previous one.

  The value of z is hard to determine. We set z equal to 0.6 because in the optimally controlled allocation problem that causes treatment's share of the budget in steady state to be around one-third, which seems like a reasonable value.

- $r = 0.04$ is the discount rate used in the previous cocaine control work (Rydell & Everingham, 1994; Caulkins et al., 1997).

Using these assumptions, the base parameter values can be computed as follows:

- From

$$c\left(\frac{u}{A+\delta}\right)^{z} A = c\left(\frac{900,000}{6,500,000+0.001}\right)^{0.6} 6,500,000 = 85,800$$

  we get $c = 0.04323$.

- The parameters d and e of the linear price function, $p(A, v) = d + e\frac{v}{A+\epsilon}$, are computed by assuming that a 1 % increase in enforcement spending causes a

$$\frac{0.2}{1 - 0.25 + (0.2 - 0)(-1)} = 0.\overline{36}$$

  % increase in price (cf. Caulkins et al., 1997). ONDCP (1997) reports a price per pure gram for purchases of 5 oz. or less of \$ 106.73 in 1992. This yields $d = 0.06792$ and $e = 0.02655$, respectively, which implies that the per gram cocaine price with minimal enforcement is around \$ 68.

- The term $\kappa A p\,(A,v)^\omega$ in the utility functional represents the social costs caused by consumption. With a total consumption of $291,000,000$ grams (Rydell & Everingham, 1994) and social costs of \$ 100 per gram consumed, the total social costs due to drug consumption amount to \$ $29,100,000,000$. We thus have

$$\kappa A\Big(d + e\frac{v}{A+\epsilon}\Big)^\omega = \kappa \cdot 6,500,000 \cdot 0.10673^{-0.5} = 29,100,000,$$

  from which we get $\kappa = 1.46259$.

- $$k\Big(d + e\frac{v}{A+\epsilon}\Big)^a = k \cdot 0.10673^{-0.25} = 1,000,000$$

  yields $k = 571,573$.

- $\mu = 0.18841$ is computed from

$$\mu\Big(d + e\frac{v}{A+\epsilon}\Big)^b A = \mu \cdot 0.10673^{0.25} \cdot 6,500,000 = 700,000.$$

A.2 How Eliminating the Per Gram Costs Parameter κ Affects the Values of d, e, k, and μ

The utility functional term $\kappa A p^\omega$, which represents the total damage that is caused by drug consumption, can be re-written in the form

$$\kappa A p^\omega = \kappa A\Big(d + e\frac{v}{A+\epsilon}\Big)^\omega = A\Big[\Big(\kappa^{\frac{1}{\omega}} d\Big) + \Big(\kappa^{\frac{1}{\omega}} e\Big)\,\frac{v}{A+\epsilon}\Big]^\omega,$$

which means that κ can be eliminated simply by transforming the price function parameters d and e by

$$d \mapsto \kappa^{\frac{1}{\omega}} d \tag{A.1}$$

and

$$e \mapsto \kappa^{\frac{1}{\omega}} e, \tag{A.2}$$

respectively. These changes, however, also cause the initiation and outflow due to enforcement terms in the state dynamics equation, kp^a and $\mu p^b A$, respectively, to change as follows:

$$kp^a \mapsto k\left[\left(\kappa^{\frac{1}{\omega}}d\right)+\left(\kappa^{\frac{1}{\omega}}e\right)\frac{v}{A+\epsilon}\right]^a=\left(\kappa^{\frac{a}{\omega}}k\right)\left(d+e\frac{v}{A+\epsilon}\right)^a=\left(\kappa^{\frac{a}{\omega}}k\right)p^a$$

and

$$\mu p^b A \mapsto \mu\left[\left(\kappa^{\frac{1}{\omega}}d\right)+\left(\kappa^{\frac{1}{\omega}}e\right)\frac{v}{A+\epsilon}\right]^b A=\left(\kappa^{\frac{b}{\omega}}\mu\right)\left(d+e\frac{v}{A+\epsilon}\right)^b A=\left(\kappa^{\frac{b}{\omega}}\mu\right)p^b A.$$

In other words,

$$k \mapsto \kappa^{\frac{a}{\omega}}k$$

and

$$\mu \mapsto \kappa^{\frac{b}{\omega}}\mu.$$

Hence, when eliminating the per gram costs parameter, we have to first change the price function parameters d and e as in (A.1) and (A.2) and second multiply k and μ by the factors $\kappa^{-\frac{a}{\omega}}$ and $\kappa^{-\frac{b}{\omega}}$, respectively, to not alter the initiation and outflow due to enforcement terms in the state equation.

A.3 Proof of the Concavity with Respect to v of the Hamiltonian in the Optimally Controlled Allocation Problem for $\lambda < 0$

The first order derivative of the Hamiltonian H in (3.6) with respect to v is given by

$$H_v=-\kappa\omega p^{\omega-1}p_v A+\lambda\left(akp^{a-1}p_v-cz\beta^{z-1}\beta_v A-\mu bp^{b-1}p_v A\right),$$

which yields

$$\begin{aligned} H_{vv} \;=\; & -\kappa\omega(\omega-1)p^{\omega-2}p_v^2A-\kappa\omega p^{\omega-1}p_{vv}A+\lambda\left[a(a-1)kp^{a-2}p_v^2+\right.\\ & +akp^{a-1}p_{vv}-cz(z-1)\beta^{z-2}\beta_v^2A-cz\beta^{z-1}\beta vvA-\\ & \left.-\mu b(b-1)p^{b-2}p_v^2A-\mu bp^{b-1}p_vvA\right] \end{aligned} \tag{A.3}$$

for the second derivative. Using β and p as in (3.2) and (3.1), respectively, we have $\beta_{vv} = p_{vv} = 0$ so that (A.3) simplifies to

$$\begin{aligned} H_{vv} &= -\kappa\omega(\omega-1)p^{\omega-2}p_v^2A + \lambda\left[a(a-1)kp^{a-2}p_v^2 - cz(z-1)\beta^{z-2}\beta_v^2A - \right. \\ &\left. -\mu b(b-1)p^{b-2}p_v^2A\right] \end{aligned}$$

which is negative if $\lambda < 0$ holds.

A.4 Derivation of the Differential Equation for the Control v in the Optimally Controlled Allocation Problem

We compute (3.10) and the expressions therein for the more general case in which the initiation term in the state equation (2.6), kp^a, is replaced by kIp^a, where $I = I(A)$ is a function that may depend on the current number of users (see Chapter 6). For this problem, the Hamiltonian and the costate variable are given by

$$H = -\kappa Ap^\omega - GA + \lambda\left(kIp^a - c\beta^zA - \mu p^bA\right)$$

and

$$\lambda = \frac{\kappa\omega p^{\omega-1}p_vA}{akIp^{a-1}p_v - cz\beta^{z-1}\beta_vA - \mu bp^{b-1}p_vA} =: \frac{Z(\lambda)}{N(\lambda)}, \tag{A.4}$$

respectively.

Differentiating λ with respect to time we get

$$\dot{\lambda} = \lambda_A\dot{A} + \lambda_v\dot{v}. \tag{A.5}$$

Equating (A.5) with the costate equation (3.9) yields

$$\dot{v} = \frac{r\lambda - H_A - \lambda_A\dot{A}}{\lambda_v}$$

where λ is given by (A.4),

$$\begin{aligned} H_A &= -\kappa p^{\omega-1}\left(p + \omega p_AA\right) - G + \lambda\left[kp^{a-1}\left(I_Ap + aIp_A\right) - \right. \\ &\left. -c\beta^{z-1}\left(z\beta_AA + \beta\right) - \mu p^{b-1}\left(bp_AA + p\right)\right] =: H_A^{(1)} + \lambda H_A^{(2)}, \end{aligned}$$

$$\begin{aligned}\lambda_A N(\lambda)^2 &= ak\kappa\omega p^{a+\omega-3}p_v^2\left[p\left(I-I_A A\right)+(\omega-a)Ip_A A\right]+c\kappa\omega z\beta^{z-2}p^{\omega-2}A^2\cdot\\ &\cdot\left\{\beta\beta_v\left[(1-\omega)p_A p_v-pp_{vA}\right]+pp_v\left[(z-1)\beta_A\beta_v+\beta\beta_{vA}\right]\right\}+\\ &+\kappa\mu b(b-\omega)\omega p^{b+\omega-3}p_A p_v^2 A^2=:Z\left(\lambda_A\right)\end{aligned}$$

and

$$\begin{aligned}\lambda_v N(\lambda)^2 &= a(\omega-a)k\kappa\omega Ip^{a+\omega-3}p_v^3 A+c\kappa\omega z\beta^{z-2}p^{\omega-2}A^2\cdot\\ &\cdot\left\{\beta_v p_v\left[(1-\omega)\beta p_v+(z-1)\beta_v p\right]+\beta p\left(\beta_{vv}p_v-\beta_v p_{vv}\right)\right\}+\\ &+\kappa\mu b(b-\omega)\omega p^{b+\omega-3}p_v^3 A^2=:Z\left(\lambda_v\right).\end{aligned}$$

In order to get the expressions for the optimally controlled allocation problem we simply set $I(A)\equiv 1$.

A.5 Derivation of the Differential Equation for the Control v in the Unrestricted Optimal Control Problem

As for (3.10) we compute the expressions in (3.13) with the initiation term being kIp^a. Then, the current value Hamiltonian is given by

$$H=-\kappa Ap^\omega-u-v+\lambda\left(kIp^a-c\beta^z A-\mu p^b A\right).$$

This implies

$$\lambda=\frac{1+\kappa\omega p^{\omega-1}p_v A}{p_v\left(akIp^{a-1}-\mu bp^{b-1}A\right)}=:\frac{Z\left(\lambda\right)}{N\left(\lambda\right)},$$

$$u=\left[\frac{cz{\beta_u}^z A\left(1+\kappa\omega p^{\omega-1}p_v A\right)}{p_v\left(\mu bp^{b-1}A-akIp^{a-1}\right)}\right]^{\frac{1}{1-z}}$$

(note that again $\beta_u=\frac{1}{A+\delta}$ does not depend on u from which we have that this expression for u depends only on A and v),

$$\begin{aligned}H_A &= -\kappa p^{\omega-1}\left(p+\omega p_A A\right)+\lambda\left[kp^{a-1}\left(I_A p+aIp_A\right)-c\beta^{z-1}\left(z\beta_A A+\beta\right)-\right.\\ &\left.-\mu p^{b-1}\left(bp_A A+p\right)\right]=:H_A^{(1)}+\lambda H_A^{(2)},\end{aligned}$$

$$\begin{aligned}\lambda_A N\left(\lambda\right)^2 &= \kappa\omega p^{\omega-3}p_A{p_v}^2 A\left[ak\left(\omega-a\right)Ip^a+\mu b\left(b-\omega\right)p^b A\right]+akp^{a-2}\cdot\\ &\cdot\left\{\kappa\omega p^\omega{p_v}^2\left(I-I_A A\right)-I_A pp_v-I\left[\left(a-1\right)p_A p_v+pp_{vA}\right]\right\}+\\ &+\mu bp^{b-2}\left[p\left(p_{vA}A+p_v\right)+\left(b-1\right)p_A p_v A\right]=:Z\left(\lambda_A\right),\end{aligned}$$

and

$$\begin{aligned}\lambda_v N(\lambda)^2 &= akIp^{a-3}\left\{\kappa\omega p^{\omega} {p_v}^3 A(\omega-a) - p\left[pp_{vv} + (a-1){p_v}^2\right]\right\} + \\ &\quad +\mu b p^{b-3} A\left\{\kappa\omega p^{\omega} {p_v}^3 A(b-\omega) + p\left[pp_{vv} + (b-1){p_v}^2\right]\right\} =: Z(\lambda_v).\end{aligned}$$

References

Baveja, A., R. Batta, J. P. Caulkins and M. H. Karwan, 1993, Modeling the Response of Illicit Drug Markets to Local Enforcement, Socio-Econ. Plann. Sci., 27(2), 73–89.

Behrens, D. A., J. P. Caulkins, G. Tragler, J. L. Haunschmied and G. Feichtinger, 1997a, A dynamic model of drug initiation: implications for treatment and drug control, Working Paper #213, Department for Operations Research and Systems Theory, Vienna University of Technology.

Behrens, D. A., J. P. Caulkins, G. Tragler and G. Feichtinger, 1997b, Controlling the U.S. Cocaine Epidemic: Prevention from Light vs. Treatment of Heavy Use, Working Paper #214, Department for Operations Research and Systems Theory, Vienna University of Technology.

Caulkins, J. P., 1993, Zero-Tolerance Policies: Do They Inhibit or Stimulate Illicit Drug Consumption?, Management Science, 39(4), 458–476.

Caulkins, J. P., C. P. Rydell, W. L. Schwabe and J. Chiesa, 1997, Mandatory Minimum Drug Sentences: Throwing Away the Key or the Taxpayers' Money? (RAND, Santa Monica).

Everingham, S. S. and C. P. Rydell, 1994, Modeling the Demand for Cocaine (RAND, Santa Monica).

Feichtinger, G. and R. F. Hartl, 1986, Optimale Kontrolle ökonomischer Prozesse – Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften (Walter de Gruyter, Berlin).

Feichtinger, G., W. Grienauer and G. Tragler, 1997, Optimal dynamic law enforcement, Working Paper (Forschungsbericht Nr. 197), Department for Operations Research and Systems Theory, Vienna University of Technology.

Johnson, R. A., D. R. Gerstein, R. Ghadialy, W. Choy and J. Gfroerer, 1996, Trends in the Incidence of Drug Use in the United States, 1919–1992 (Substance Abuse and Mental Health Services Administration, Office of Applied Studies, U.S. Department of Health and Human Services).

Kleiman, M. A. R., 1988, Crackdowns: The Effects of Intensive Enforcement on Retail Heroin Dealing, in Chaiken, M. R. (ed.), Street-Level Drug Enforcement: Examining the Issues (National Institute of Justice, Washington, D.C.).

Kleiman, M. A. R., 1993, Enforcement Swamping: a Positive-Feedback Mechanism in Rates of Illicit Activity, Mathl. Comput. Modelling, 17(2), 65–75.

Kort, P. M., G. Feichtinger, R. F. Hartl and J. L. Haunschmied, 1996, Optimal Enforcement Policies (Crackdowns) on a Drug Market, Discussion Paper No. 9629, CentER for Economic Research, Tilburg University.

Leonard, D. and N. V. Long, 1992, Optimal control theory and static optimization in economics (Cambridge University Press, Cambridge).

Miller, T. R., M. A. Cohen and B. Wiersema, 1996, Final Summary Report presented to the National Institute of Justice.

Moore, M. H., 1973, Policies to Achieve Discrimination on the Effective Price of Heroin, American Economic Review, 63, 270-279.

Musto, D. F., 1987, The American Disease: Origins of Narcotic Control (Oxford University Press, New York).

Office of National Drug Control Policy, 1997, The National Drug Control Strategy, 1997: Budget Summary (The White House, Washington, DC).

Office of National Drug Control Strategy, 1996, The national drug control strategy: 1996 (The White House, Washington, DC).

Reuter, P. and M. A. R. Kleiman, 1986, Risks and Prices: An Economic Analysis of Drug Enforcement, in Tonry, M. and N. Morris (eds.), Crime and Justice: An Annual Review of Research, Vol. 7 (University of Chicago Press, Chicago), 289–340.

Rhodes, W. P. Scheiman, T. Pittayathikhun, L. Collins and V. Tsarfaty, 1995, What America's Users Spend on Illegal Drugs, 1988–1993 (Office of National Drug Control Policy, Washington, DC).

Rice, D. P., S. Kelman, L. S. Miller and S. Dunmeyer, 1990, The Economic Costs of Alcohol and Drug Abuse and Mental Illness: 1985, report submitted to the Office of Financing and Coverage Policy of the Alcohol, Drug Abuse, and Mental Health Administration, U.S. Department of Health and Human Services, Institute for Health and Aging, University of California, San Francisco, CA.

Rydell, C. P., J. P. Caulkins and S. S. Everingham, 1996, Enforcement or treatment? Modeling the relative efficacy of alternatives for controlling cocaine, Operations Research, 44(6), 687–695.

Rydell, C. P. and S. S. Everingham, 1994, Controlling cocaine – supply versus demand programs (RAND, Santa Monica).

Skiba, A. K., 1978, Optimal growth with a convex-concave production function, Econometrica, 46, 527–539.

Varian, H. R., 1996, Intermediate Microeconomics: A Modern Approach, 4th edition (W. W. Norton & Company, New York).

Wolfram, S., 1996, The Mathematica Book, Third Edition (Cambridge University Press, Cambridge).

Glossary of Variables

a	elasticity of initiation with respect to price
$A(t)$	number of users at time t
$A_0 = A(0)$	initial number of users
A_{min}	minimum level of users along the left stable manifold of the high equilibrium in the optimal control problem with state dependent initiation term
A_S	critical level (*Skiba* threshold)
$\underline{A}$	minimum steady state value of A in the optimal control problem with state dependent initiation term
$_a$	index associated with variables of the optimally controlled **a**llocation problem
$_{ai}$	index associated with variables of the optimally controlled **a**llocation problem with state dependent **i**nitiation term
$_A$	index to indicate derivatives with respect to A
α	exponent in the state dependent initiation factor $I(A) = A^{\alpha}$
b	elasticity of quitting with respect to price
$B(t)$	total budget spending at time t
$\beta\left(\frac{u(t)}{A(t)+\delta}\right)$	efficiency of treatment at time t
c	efficiency of treatment
$C(t)$	total social costs due to consumption at time t
$_c$	index associated with variables of the **c**onstant fraction control model
$d = p\,(v(t) = 0)$	price with minimal enforcement

δ	constant to avoid division by zero
$\dot{}$	dots indicate derivatives with respect to time
e	efficiency of enforcement
$e_{\dots}$	elasticities
ϵ	constant to avoid division by zero
$f(t)$	enforcement's share of total budget at time t
$_f$	index associated with variables of the unrestricted (**f**ree) control problem
$_{fi}$	index associated with variables of the unrestricted (**f**ree) control problem with state dependent **i**nitiation term
G	drug control budget (in thousands of dollars) per addict
G_{pgc}	value of G, for which the utility functional is maximized, if *p*er *g*ram *c*osts are at \$ *pgc*
H	current value Hamiltonian
$^{(h)}$	superindex associated with variables of the **h**igh equilibrium in the optimal control problem with state dependent initiation term
$\hat{}$	hats are associated with steady state values
$I(A(t))$	initiation depending on the current number of users
J	utility functional
$\underline{J}^*$	optimal utility functional value along the trajectory from $\left(\hat{A}^{(l)}_{fi}, \hat{v}^{(l)}_{fi}\right)$ to $(\underline{A}, \underline{v})$ in the optimal control problem with state dependent initiation term
$J^{(h)*}$	optimal utility functional value along the trajectory from $\left(\hat{A}^{(l)}_{fi}, \hat{v}^{(l)}_{fi}\right)$ to $\left(\hat{A}^{(h)}_{fi}, \hat{v}^{(h)}_{fi}\right)$ in the optimal control problem with state dependent initiation term
k	initiation proportionality constant (new users / year)
k_{base}	base value of k
k_{high}	initiation proportionality constant in the first (finite time) period of the two phases problem
k_{low}	initiation proportionality constant in the second (infinite time) period of the two phases problem

κ	per gram costs proportionality constant
κ_{pgc}	indicates that *p*er *g*ram *c*osts are at \$ *pgc*
$^{(l)}$	superindex associated with variables of the **l**ow equilibrium in the optimal control problem with state dependent initiation term
$\lambda(t)$	current value costate variable
μ	outflow rate from use
$N(\lambda)$	denominator of the expression for λ
$_n$	index associated with variables of the uncontrolled (**n**o control) model
ω	short run elasticity of demand
$p\left(\frac{v(t)}{A(t)+\epsilon}\right)$	drug price at time t
p_{base}	base price of p
r	annual discount rate
$S(A(T))$	value of being in the state $A(T)$ at the end of the planning horizon (*salvage value function*)
*	stars are associated with optimal values
t	time
T	duration of the high k period in the two phases problem
τ	time when the government starts to control or time which is needed to move from A_S to $\underline{A}$
$u(t)$	budget for treatment spending at time t
$v(t)$	budget for enforcement spending at time t
$\underline{v}$	minimum steady state value of v in the optimal control problem with state dependent initiation term
$_v$	index to indicate derivatives with respect to v
z	$1-z$ reflects extent of diminishing returns to treatment
$Z(\lambda)$	numerator of the expression for λ
$Z(\lambda_A)$	numerator of the expression for λ_A
$Z(\lambda_v)$	numerator of the expression for λ_v

List of Figures

List of Tables