ESSAYS IN MATHEMATICAL ECONOMICS AND FINANCE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT
SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bruno Strulovici
August 2005

UMI Number: 3187348

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Thomas A. Weber Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

J. Darrell Duffie
(Graduate School of Business)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Paul R. Milgrom
(Department of Economics)

Approved for the University Committee on Graduate Studies.

Preface

This thesis consists of three theoretical essays in economics and finance. A common characteristic of these essays is their emphasis on mathematical methods to generalize existing results in their respective areas, which are 1. Monotone comparative statics [MCS], 2. Games played through agents [GPTA], and 3. Performance-sensitive debt [PSD]. In all essays, agents are utility or profit maximizers. MCS considers a simple economic setting in which an agent maximizes a parameterized objective function with respect to a constrained decision vector. The aim of the essay is to find robust relations between parameters and optimal decision variables. In GPTA, decision makers are divided into two types: principals and agents. Principals move first, proposing monetary contracts to agents, who then take actions to maximize their respective utility. The essay is primarily concerned with the construction and analysis of contracts that implement socially efficient outcomes. PSD considers a firm whose shareholders carry the common objective of maximizing the firm's equity value while raising a given amount of cash, through the issuance of a *performance-sensitive debt* contract, whose interest payments depend on some performance measure of the firm.[1] The shareholders' decision variable is a stopping time, at which the firm declares bankruptcy, and which is affected by the particular debt contract that shareholders choose to issue.

The economic questions analyzed in the essays are introduced in Chapter 1, which also summarizes my contributions to these questions. Because of the mathematical emphasis of the thesis, Chapter 1 also contains a presentation of the quantitative

[1] Contracts with fixed interest rate are considered performance-sensitive, with a trivial, null sensitivity.

methods that have proved useful in my research.

Chapter 2 is an adaptation of the paper "Monotone Comparative Statics: A Geometric Approach," which was co-written with my principal adviser, Thomas Weber. When we started this project, Thomas proposed an analogy with the "Reynolds Number": in the same way as this number, in fluid mechanics, is all the information needed from a system in order to determine its turbulence behavior, we would look for a number summarizing all relevant information about the parameters of a problem, to describe movements in the optimal decision variables resulting from changes in the parameters. To guide us, we also had an example[2] where decision variables and parameters had been transformed to obtain monotone comparative statics: in a multi-period production planning problem with convex production and storage costs, the optimal *cumulative* production vector was nonincreasing in the *cumulative* sales vector.[3] In this example, the reason to consider cumulative variables, as opposed to more natural period variables, precisely stems from the fact that monotone comparative statics only obtain with the former. Using cumulative variables also permits to model any sales delay as a decrease in the vector of cumulative sales, and thus obtain the intuitive fact that a sales delay reduces the optimal vector of cumulative production. In this example, therefore, the transformation of parameters and variables was instrumental in obtaining monotone comparative statics. For us, the challenge was to find a systematic way to obtain such transformations of the parameters in *any* problem. The first essay of this thesis proposes a way to take it up.

"Constructing Efficient Equilibria in Games Played Through Agents," a paper also co-written with Thomas Weber, constitutes the third chapter of the thesis. I first got involved in this project when Thomas was working with another doctoral student, Hongxia Xiong,[4] on a very similar problem in the context of two-echelon supply chains. In the operations management literature, a supply chain is said to be "coordinated" if the firms in the supply chain take actions that maximize the sum of their payoffs, as if

[2]This example was suggested to us by Arthur F. Veinott, Jr. and first introduced to me in his class on Supply Chain Optimization, in Fall 2000.

[3]In this problem, demand at each period is assumed to be deterministic, and must be perfectly met by the scheduled production.

[4]See Weber and Xiong (2004).

a planner were coordinating the different firms to achieve this optimal outcome. The supply chain literature contains many instances of parametric contracts coordinating supply chains in particular contexts. Prat and Rustichini (2003) offered a basis to unify and generalize these results, which have many economic applications beyond the context of supply chains. The essay proposes a conceptual framework to think about efficient contracts in games with multiple principals and multiple agents, and provides explicit constructions to obtain a general set of such contracts. Moreover, it characterizes, within this set, the contracts that are most beneficial for the principals, and thus most likely to be chosen by them.

Finally, Chapter 4 is based on the paper "Performance-Sensitive Debt," which was co-written with two other doctoral students, Alexei Tchistyi and Gustavo Manso, and supervised by Darrell Duffie. At the time we started this project, in January 2002, step-up bonds, the main protagonist in our class of performance-sensitive debt contracts) had experienced a surge of popularity. In addition to their volume increase, step-up bonds also starred in several major financial magazines. By the completion of our paper, however, the step-up bond craze had faded.[5] This is comforting: the main theorem of our paper shows that, compared to fixed-coupon bonds, step-up bonds are socially inefficient.[6]

[5]In fact, the popularity of *rating-triggered* step-up bonds, which we investigate, has been so much surpassed by other bonds, that the expression "step-up bond" has come to designate a different kind of bonds.

[6]This result calls for agency-cost and renegotiation-cost justifications for the existence of performance-sensitive debt contracts. In fact, performance-pricing loans are still popular, especially among syndicated loans, where these costs are particularly salient (see Asquith, Beatty and Weber (2002)).

Acknowledgements

This thesis owes to many people. Among them, I am extremely grateful to my principal adviser, Thomas Weber, for his exceptional guidance throughout my last year at Stanford. Thomas has created the means to conduct exciting and rigorous research, which was the greatest gift that I received at Stanford. My interest in finance was developed during Darrell Duffie's remarkable class, Dynamic Asset Pricing Theory. Darrell has distilled a remote guidance during almost the whole duration of my doctoral studies, and got me started on the step-up bond project, which has proved an invaluable experience. During this project, I learned much through endless conversations with my fellow doctoral students, Gustavo Manso and Alexei Tchistyi. During Fall 2003, I frequently met Anat Admati, whose sharp and insightful comments helped me understand the importance of economic assumptions in financial models.

I am very grateful to Jim Primbs, who supported the agent-based simulation project which I developed during my third year. Jim also helped me write my first papers, took me to my first conference, and helped me lead the computational team of the project. His open-mindedness and friendliness allowed me to keep a free spirit and sense of initiative, and follow my intuition and ideas.

Over the last six months, I had the privilege to interact with Paul Milgrom, who gave me a chance to prove myself at the most important time, and put me in the middle of the microeconomic action, by sharing with me some of his on-going research questions to work on. I am only one among many for whom Paul's work in economics serves as a source of inspiration and motivation. This is also thanks to Paul's suggestion that I could discover my new job as a Nuffield Postdoctoral Research Fellow at Oxford University.

I benefited from many excellent classes at Stanford. Among them, I was particularly struck by Douglas Bernheim and Antonio Rangel's class on psychology and economics. This class offered highly insightful discussions on experimental economics and its relation to the standard economic model. I acquired fundamental knowledge from excellent classes, including Ken Singleton's Empirical Asset Pricing, Ilya Segal's Contract Theory, Jeff Strnad's Tax and Finance, Pete Veinott's Supply Chain Optimization, George Papanicolaou's Stochastic Differential Equations, David Luenberger's Vector Space Optimization, and Peter Glynn's doctoral classes.

In terms of guidance, I am also very grateful to my friends Soong Moon Kang and Nicolas Méary, whose experience at Stanford and in life, openness, and sharp spirit were very helpful more than once.

My life at Stanford would not have been as fun without my incredible office mates Yen Lin Chia, Chris Messer and Sachin Jain who made sure I was never bored in and out of the office, and my other numerous friends, Jose and Lalli Blanchet, and Boris and Yue Postler, to mention a few. My background in mathematics, which was my main asset at Stanford (along with perseverance!), was built in France. It owes a lot to my father, who inculcated me with the sense of rigor and logic, and to my mathematics professors Bernard Lanoizelée and Michel Lambert.

Last, I am very grateful to my family, especially my sister Isabelle, for their support and affection. I have no words to express my thanks to my parents and Hiroko for the love they give me every single day.

Contents

Chapter 1

Introduction

1.1 Motivations

The essays address several types of economic questions. GPTA focuses on social efficiency of contracts: given agents' action sets and players' payoff functions, can principals propose contracts[1] inducing actions that maximize the total surplus of the economy? Bernheim and Whinston (1986) study the case of multiple principals and a single agent. In order to prove the existence of socially efficient contracts, they introduce a refinement of subgame perfect Nash equilibrium (the game has two stages, with principals moving first, and the agent moving second), called *truthful Nash equilibrium.* In this refinement, the contract between any principal and the agent perfectly reflects the impact on the principal's payoff of the agent's possible deviation. For example, suppose that some principal P gets a gross payoff of 100 and pays 10 to the agent at equilibrium. Then, in a truthful Nash equilibrium, P's

[1] By definition, a contract between a principal and an agent is a function of the agent's action, specifying the monetary transfer from the principal to the agent for each possible action in the agent's action set.

contract will compensate any deviation by the agent causing his payoff to shift to $100 + \Delta$ by a modified transfer of $10 + \Delta$ to the agent, resulting in P's net payoff to be unaffected by A's deviation. Bernheim and Whinston show that – under general conditions – any efficient outcome can be implemented as a truthful Nash equilibrium.[2] In numerous applications, there are several agents. For example, lobbies (principals) influence multiple decision makers (agents) to weigh on political decisions. In multi-unit auctions, bidders (principals) sometimes face several auctioneers or owners (agents). In the context of supply chains, producers (principals) enter advertising contracts with multiple retailers (agents) to promote their products. In the insurance industry, insurance companies (the principals) propose compensation schemes to insurance brokers (the agents). These examples raise the following questions: does Bernheim and Whinston's existence result generalize to the multiple-agent setting? For practical applications, is it possible to go beyond the existence result, and propose explicit contracts implementing a socially efficient outcome? In general, there may exist several equilibria whose contracts implement a given efficient outcome. Is there a solution concept that reduces the set of such contracts for predictive and normative purposes? If such refinement exists, can we characterize contracts that satisfy the corresponding additional requirements?

PSD considers social efficiency in a different context: among the class of performance-sensitive debt contracts[3] raising a given amount of cash for a firm, which contracts minimize expected bankruptcy cost? In our model, this minimization is equivalent to equity maximization. The question of social efficiency is therefore closely related to shareholders' optimal decision problem. Choosing which debt contract to issue has

[2] Since an efficient outcomes is by definition a maximizer of the total surplus, there can exist several of them.

[3] As defined in the Preface, these are contracts whose interest rate depends on some performance measure of the issuer.

become a more salient issue in the past decade, which has experienced a popularization of contracts bearing various covenants with potential influence on firms' default behavior.[4] Since the type of debt contract chosen may influence default, how should investors value performance-sensitive debt contracts? A first difficulty stems from the particularity that coupon payments are stochastic. The type of stochasticity considered here is of another nature than floating-rate debt contracts: the former is based on credit risk, while the latter is based on interest rate risk. Secondly, payments are interrupted whenever the firm defaults, which clearly affects the value of debt.[5] Since default is determined endogenously by shareholders, whose behavior depends on the type of debt issued, investors should discount the effect of performance-sensitivity of debt in their valuation. Moreover, performance measures often depend on a third party's perception of the issuer's performance (such as the credit note delivered by a rating agency). How does performance sensitivity of debt affect this perception, and how does this perception affect 1. shareholders' default behavior after a debt contract has been issued, and 2. shareholders' choice of debt contract?

MCS studies a cross-sectional issue in economics: in a parameterized problem, how do parameter moves affect agent decisions? The concept of *comparative statics*[6] was introduced by Samuelson (1941) to formalize this ancient question. A classic example is the impact of distortionary tax on the total surplus of an economy. It is well-known that when tax is added to the market price faced by consumers, the reduction of their demand results in a surplus loss (even when the proceeds of the tax are given back to consumers), graphically represented as the celebrated deadweight loss

[4]See Section 4.2 for examples.

[5]While interest risk clearly plays a role on the value of debt contracts, we focus on credit risk and assume constant interest rate, in the spirit of Leland (1994).

[6]For simplicity, we adopt the usual convention that "comparative statics" stands for *monotone* comparative statics, meaning there is a stable relation between optimal decision variables and parameters.

triangle. This result can be proved analytically using the implicit function theorem.[7] Unfortunately, in a more complex setting, it might not be possible to obtain such simple comparative statics. For example, Giffen and Veblen[8] goods generate consumer (uncompensated) demand that is not monotonic in their price. When comparative statics fail to obtain with the original parameters of the problem, is there a way to construct a function of the parameters, an "index", such that variations of optimal decision variables following changes in the parameters can be monotonically described by the index? How does one interpret such index? What type of economic questions can be addressed using comparative statics with transformed parameters?

1.2 Contributions

GPTA is based on Prat and Rustichini (2003), who introduce the concept of weakly truthful equilibrium to give a nonconstructive proof (based on the notion of "balancedness") for the existence of efficient contracts[9] in their game of complete information with multiple principals and multiple agents. The solution concept of weakly truthful equilibrium is a refinement of subgame perfect Nash equilibrium but, as the name indicates, weaker than the truthful Nash equilibrium introduced by Bernheim and Whinston. The approach presented in GPTA goes further by explicitly constructing efficient contracts, which proves *de facto* their existence. We show that in the generic case, efficient contracts can be determined by simple differentiation of principals' payoff functions.[10] In order to further investigate efficient contracts, we introduce

[7]See for example Mas–Colell, Whinston, and Green (1995), pp. 331–332.
[8]See Section 2.5.3.
[9]Throughout, "efficient contracts" means contracts that implement an efficient outcome of the game.
[10]We assume, like Prat and Rustichini, that payoff functions are concave.

concepts that clarify the mathematical structure of weakly truthful equilibria, allowing us to construct a general – under some conditions, exhaustive – set of weakly truthful equilibria implementing any given efficient outcome. GPTA also investigates the allocation of total surplus between principals and agents. Since principals move first and possess all bargaining power, it is reasonable to predict that they will (or assert that they should) try to extract the highest possible fraction of the surplus. We refine the set of weakly truthful equilibria constructed earlier by characterizing the subset of contracts that are Pareto optimal for the principals.

PSD is based on the endogenous-default model of Leland (1994), which it generalizes in two ways. First, while Leland focuses on fixed-coupon bonds to derive an important formula for the optimal default triggering level of the firm, our approach allows for any performance-sensitive (including discontinuous) debt profiles. Such profiles include step-up bonds,[11] performance-pricing loans, and other examples described in the essay. This generalization allows us to compare the efficiency of different debt contracts in terms of bankruptcy cost. Specifically, we introduce a partial order within the class of performance-sensitive debt contracts, and show that this partial order is stronger[12] than the efficiency order, where debt contracts are ranked according to their implied expected bankruptcy cost. In order to be carried out, the analysis requires consideration of both endogenous default and performance-sensitivity of debt, which the essay is the first to explicitly model together. The relation between the two orders implies that *risk-compensating* debt contracts, whose interest payments increase with credit risk, are inefficient compared to contracts with

[11]Throughout, "step-up bonds" stands for bonds whose coupon rate at any time is a function of the issuer's credit rating (e.g. the credit note given by Moody's or Standard & Poor's rating agencies), with the coupon rate increasing as the rating deteriorates. This definition should not be confused with another sense of "step-up bonds", in which the coupon rate increases according to a predetermined schedule. With this second sense, the coupon rate is a deterministic function of time.

[12]Order $>$ is stronger than order $\succ$ if $x > y$ implies $x \succ y$ for any x, y in the decision set.

fixed interest rate. Second, the ordinary differential equation that Leland uses to compute the value of the equity and optimal default triggering level is generalized to the case where the asset level follows any diffusion process (up to the usual assumptions for the existence of a strong solution to the stochastic differential equation defining its dynamics). Moreover, the inefficiency result of the essay is shown to hold under these much weaker conditions. We then consider the fact that performance measures often depend on a third party's perception of the issuer's performance (such as the credit note delivered by a rating agency). We formalize the circularity issue arising from the reciprocal influence between shareholders' and the third party's behaviors. We resolve it in the case of step-up bonds, where asset thresholds defining credit ratings reflect the firm's actual probabilities of default. We draw implications of our model for the behavior of credit-rating agencies.

The field of monotone comparative statics has experienced a revolution with[13] the introduction of the concept of supermodularity. The supermodularity assumption (and the related "quasisupermodularity" and "single-crossing" properties) plays a similar role in comparative statics as convexity in optimization theory: it drastically simplifies the theory, and yields very strong results, as Milgrom and Shannon (1994) illustrate. On the other hand, it also rules out many interesting problems: restricting comparative statics to supermodularity is analogous to restricting optimization theory to convex objective functions. Much research so far has focused on developing sophisticated techniques based on supermodularity to derive comparative statics. The implicit function theorem (IFT) has been used to show comparative statics before the concept of supermodularity even existed, but only to solve ad hoc problems. MCS restores its status to this theorem, making it the cornerstone of a systematic

[13]See Veinott (1965) and Topkis (1968).

way to derive comparative statics. When a problem involves multiple parameters, we create paths in the parameter space – whenever our information about the structure of the problem is precise enough – along which optimal decision variables (or a subset, or a function, thereof) are monotonic. We then show how these paths can generate a new parametrization of the problem. Under some differentiability and regularity assumptions, we also establish the connection between our approach and supermodularity-based results. We show that the joint condition of supermodularity among decision variables and between decision variables and parameters implies that optimal decision variables are nondecreasing along any path pointing "north-east" in the parameter space.[14] We illustrate the potential of the method though several applications. First of all, the approach allows one to deal with examples where classic results simply do not hold: there are no initial parameters in which decision variables are monotonic. In this case, the approach *constructs* new parameters, an undertaking that is clearly beyond the scope of the supermodularity approach. Second, even when classic MCS results hold, it proposes a systematic way to discover them. To get their full force, the supermodularity approach requires significant expertise in its numerous techniques. By contrast, our method only requires differentiation and inversion of a matrix, to compute the pseudo-gradient introduced in MCS. If optimal decision variables are monotonic in the parameters, all pseudo-gradients will *necessarily* be nonnegative, which will be usually be clear from a simple inspection of their formula. Third, the supermodularity approach assumes that a constraint set is a lattice, a restriction that rules out, for example, any budget constraint in an economy with

[14]However, this result is less powerful than the most advanced results on supermodularity, for two reasons. First, it requires differentiability and regularity assumptions, which are unnecessary under the supermodularity-based approach. Second, it requires supermodularity (or any increasing transformation thereof), instead of the weaker quasisupermodularity and single-crossing property required by Milgrom and Shannon (1994).

more than two goods.[15] By contrast, our method, combined with the Kuhn-Tucker theorem, allows for general constraint sets, such as budget constraints.

1.3 Quantitative Methods

The essays offer several simple closed-form formulas, which can be easily used in applications, and more complex algorithms. PSD proposes closed-form formulas to price step-up bonds with any finite number of steps, whenever the asset level follows a geometric Brownian motion. The formula involves the computation of the equity value of the firm as function of its asset level. The equity value is shown to satisfy a second-order ordinary differential equation (ODE) with coefficients that are constant on any interval where the coupon rate is constant and whose boundary conditions are partly obtained by imposing the celebrated "smooth-pasting" condition on the default level. Jumps in the coupon rates are included by solving the ODE piecewise on each interval where the coupon is constant, and imposing that the equity value be continuously differentiable everywhere (including where jumps occur), which allows the determination of integration constants. Continuous differentiability is a distinct requirement from smooth-pasting, and is imposed to allow the application of Itô's formula to verify optimality.[16] When the asset level follows a general diffusion process and the coupon payment is any function of the performance measure with a locally finite number of jumps, obtaining an ODE for the equity value is more difficult.[17] To

[15] In a two-good economy, there is an astute way to get rid of the problem, by considering a particular order on planar vectors.

[16] Itô's formula can be applied to any function that is continuously differentiable and has a second order derivative with a locally finite number of discontinuities. See Karatzas and Shreve (1991).

[17] Harrison (1985) treats the case of Brownian motion and, by an easy extension, the case of geometric Brownian motion. However, the monograph does not consider general diffusions in an optimal stopping problem.

understand why, observe that the equity value[18] takes the form

$$W(x) = E_{X_0=x}\left[\int_0^{\infty} e^{-rt} f(X_t)dt\right],$$

where $\{X_t\}_{t\geq 0}$ is a diffusion and f is some function with a locally finite number of jumps, satisfying integrability conditions. To compute $W(x)$, the first step is to remark that, under some technical assumption, $W(x)$ can be rewritten as

$$W(x) = \int_0^{\infty} e^{-rt} E_{X_0=x}\left[f(X_t)\right] dt.$$

Therefore, we have simplified the computation of $W(x)$ to the evaluation of the integrand

$$E_{X_0=x}\left[f(X_t)\right].$$

This integrand, however, looks deceptively simple. In the case of a general diffusion, its computation involves the entire functions $\mu(\cdot)$ and $\sigma(\cdot)$ defining the dynamic of the process $\{X_t\}_{t\geq 0}$. We circumvent this problem by deriving an ordinary differential equation for W, instead of trying to compute it directly (as would be possible with geometric Brownian motion). A first approach consists in using Malliavin calculus,[19] to compute $W'(x)$ and $W''(x)$. However, this approach, relatively simple in the case of geometric Brownian motion, becomes intractable with general diffusions. PSD uses instead the concept of fundamental solution of a diffusion process, as analyzed by Friedman (1974). The fundamental solution attached to a diffusion process directly connects stochastic differential equations and partial differential equations, allowing

[18]We assume for the moment that default does not occur. Optimal default is considered separately, see the end of this section.

[19]See the excellent papers by Fournié et al. (1999) and (2001) in the context of financial engineering, and Nualart (1995) for a mathematical treatment.

one to analyze quantities like the integrand $E_{X_0=x}[f(X_t)]$. The approach is illustrated in Section 4.4.1. The literature on stochastic calculus contains similar optimal stopping problems. For example, Karatzas (1984), Section 3, studies an optimal stopping problem with continuous payoff (whereas we need discontinuous payoffs to account for step-up bonds). His strategy is to solve a variational inequality, involving an ODE, and use classic ODE techniques (based on the Wronskian) to express the solution of the optimal stopping problem as a function of homogeneous and particular solutions of the ODE, then verify optimality. By contrast, the strategy used in PSD is to first express the optimal solution as a function[20] of the expected reward if the process is never stopped, using the strong Markov property, then show that this latter function satisfies an ODE on any continuity interval of the payoff function. With this second approach, the (piecewise) ODE appears as a necessary condition, rather than an educated guess for sufficiency. In particular, the simplicity of the approach allows to deal with discontinuities in the payoff function, an undertaking that seems more complex with Karatzas' approach, whose proofs are already significantly more tedious. The approach of the essay is new to my knowledge, and the corresponding optimal stopping problem does not seem to appear in the finance literature at that level of generality. The approach has potential applications in other optimal stopping problems involving more general processes than geometric Brownian motion, and general payoff functions.

MCS uses the implicit function theorem (IFT) in a novel way: when the optimizer's location in the constraint set is only partially known (or even unknown), the IFT is used to identify potential directions, called *pseudo-gradients* of monotonicity. These potential directions are then used to extract – whenever possible – a vector

[20]Optimality is characterized by the "smooth-pasting condition," which appears naturally in my approach, as an immediate consequence of the envelope theorem.

field of the parameter space, such that at each parameter, the corresponding element of the vector field makes non-negative scalar product with all pseudo-gradients corresponding to the parameter. Trajectories of the vector field are paths of the parameter space, such that comparative statics obtain along these paths. We can connect the approach to the supermodularity approach. When the objective function is supermodular or any increasing function of a supermodular function (which constitutes a much larger subset of the class of quasisupermodular functions, see Milgrom and Shannon (1994)) in the decision variables and in each pair consisting of one decision variable and one parameter, all pseudo-gradients are nonnegative, from which it is immediate that any vector field consisting of nonnegative vectors makes nonnegative scalar product with the pseudo-gradient. This implies that optimal decision variables are nondecreasing along any path always pointing in the positive orthant of the parameter space and, therefore, nondecreasing in the parameters. In that sense, our approach encompasses the supermodularity-based approach. In general, it is often possible to integrate the paths of the vector field to obtain a new parametrization of the problem, where some of the new parameters are indices with respect to which comparative statics can be expressed. In order to deal with constrained optimization problems, the IFT can be used in conjunction with the Kuhn-Tucker theorem (see for example Bertsekas, 1995, p. 255). This combined approach provides a systematic way (the first, to my knowledge) to analyze situations where parameters affect not only the objective function, but also the constraint set of the optimization problem. The essay provides an example of this type.

In GPTA, the proof of the existence of socially efficient contracts is based on a characterization of weakly truthful equilibria due to Prat and Rustichini (2003) and refined by Weber and Xiong (2004), in terms of a system of inequalities whose variables of the contracts' excess payment functions (where excess is understood with respect to the efficient outcome). The essay presents an algorithm to derive a solution of this system of inequalities based on an iterative application of the separating hyperplane theorem.[21] The separating hyperplane theorem allows one to find excess payment functions, principal by principal. Construction of a general class of contracts implementing any given efficient outcome is made possible by a *leveling algorithm*, which generates minimal additive functions above the excess payoff function of each principal. The essay provides conditions – based on submodularity – under which the minimal functions are all equal, in which case we call them the *additive upper envelope* (at a given point) of the excess payoff function. Pareto optimality is characterized through an *ironing algorithm.*[22]

In the context of this game, where principals' decision variables are entire functions (as opposed to mere vectors), focusing on the maximization of a given principal's payoff is usually not enough to obtain a Pareto optimal equilibrium. In fact, there are in general many contracts maximizing both total surplus and a given principal's payoff. The ironing algorithm is crucial to isolate contracts that "care" just enough about the given principal to achieve her optimal payoff, but otherwise distribute to other principals all remaining slack in agents' utilities.[23]

[21]For this algorithm to work, we need to impose that all payoff functions are concave, which Prat and Rustichini also assume in their nonconstructive proof.

[22]In economics, "ironing" usually refers to the procedure introduced by Mussa and Rosen (1978) in the context of screening. However, this procedure and the algorithm of the essay have no connections (apart from the visual effect of their application that motivates the name).

[23]When the utility of a given consumer is increasing with respect to all goods, maximizing this consumer's utility leads to a Pareto optimal allocation: it is impossible to increase other consumers' utilities without reducing that of the singled-out consumer. The situation in games played though agents is more complex, as principals only care about the minimum of their excess payment functions, which are the variables of the efficiency problem (see Section 3.5).

Chapter 2

Monotone Comparative Statics

We consider comparative statics of solutions to parameterized optimization problems. A geometric method is developed for finding a vector field that, at each point in the parameter space, indicates a direction in which monotone comparative statics obtain. Given such a vector field, we provide sufficient conditions under which the problem can be reparameterized on the parameter space (or a subset thereof) in a way that guarantees monotone comparative statics. A key feature of our method is that it does not require the feasible set to be a lattice and works in the absence of the standard quasi-supermodularity and single-crossing assumptions on the objective function. We illustrate our approach with a variety of applications.

2.1 Introduction

In many problems of economics, important insights can be derived from a formal model by comparing its predictions for different parameter values. The model's *parameters* are *exogenously* specified and can often be varied for analysis purposes while its *variables*, which constitute the building blocks for its predictions, are *endogenously* determined by (i) imposed model relations and (ii) parameter values. For instance, an economic model might be concerned with a firm's optimal production of widgets (e.g., in terms of capital and labor requirements), given both a production function relating output to inputs and a set of prices (e.g., the market price for widgets, the cost of capital, and an average wage rate). The model would then in the neoclassical tradition impose maximization of the firm's profit to determine optimal amounts of factor inputs as a function of their respective prices. More generally, if a model's predictions can be expressed as an optimal action in some finite-dimensional space, then *comparative statics* studies the direction in which the optimal action changes consequent upon some disturbance in the values of the model's parameters (Samuelson, 1941).[1] Thus, in our example, normalizing the price of the firm's output to one, the optimal choice of inputs critically depends on the prices of the production factors which are this model's parameters.

The key question of *monotone comparative statics* is to determine under what conditions the model predictions vary monotonically with the parameters (Topkis 1968, 1998). A general answer to this question for optimal actions chosen from feasible sets, which are usually assumed to be lattices satisfying a set-monotonicity requirement with respect to the parameters, is provided by Milgrom and Shannon (1994). They provide a necessary and sufficient condition for optimal actions to exhibit monotone comparative statics with respect to the parameters. For the special case of our neoclassical production decision problem, where the production function is independent of the factor prices, monotone comparative statics (i.e., inputs nonincreasing in prices) obtain if and only if the production function is supermodular, which – assuming twice

[1] If an optimal action is infinite-dimensional (e.g., the solution *function* of a variational problem), then comparing the model's predictions for different parameter values is often referred to as "comparative dynamics." We limit our attention here to the finite-dimensional case corresponding to "comparative statics."

continuous differentiability – amounts to requiring that all cross-partial derivatives of the production function are nonnegative.

The practical importance of monotone comparative statics, justifying its widespread use in economics, lies in the fact that robust insights can be obtained in the absence of an analytical solution to the model: the monotonicity of optimal actions in parameters is guaranteed if the system's objective function satisfies certain easy-to-check requirements. In addition, monotonicity of optimal actions can yield useful rules of thumb for decision makers and thus help in arriving at "optimally imperfect decisions" (Baumol and Quandt, 1964). Clearly, in our production example (which is examined more closely in Section 2.5.2) it would be helpful for the firm to be able to immediately (i.e., without any further computations) translate price movements into appropriate input changes which at the very least vary in the right *direction*, even when its production function is not supermodular due to anticomplementarities between factor inputs. Unfortunately, the currently available theory on monotone comparative statics returns negative results in situations where the aforementioned characterization of monotone comparative statics by Milgrom and Shannon fails. We argue that this failure is often due to the fact that the parameterization of the problem is taken as given.[2] Indeed, our results indicate that it may be possible to achieve monotonicity of solutions in new parameters that are obtained by a one-to-one mapping from the original parameter space. In fact, for problems with a smooth analytical structure we show that, provided sufficiently precise knowledge about the location of an optimal action in the action space, *it is always possible to find a reparameterization that achieves monotone comparative statics in any single component of the decision.*

[2]This is true in the literature, except for some rare cases where trivial reparameterizations such as a change of sign or other simple *ad hoc* reparameterizations are chosen under very special circumstances, e.g., by Granot and Veinott (1986) in a network flow problem.

Our central goal is to provide a new method for achieving monotone comparative statics of solutions to parameterized optimization problems, first by relaxing and thereby generalizing the standard monotone comparative statics problem, and second, by providing a systematic way to reparameterize the problem in such a way that monotone comparative statics can be achieved.[3] In addition to addressing cases where the standard results do not apply, our method can, by offering a new description of the parameter space, shed light on important relations between decision variables and parameters of economic problems.

The approach developed here builds on tools in differential geometry and we thus require models with a smooth structure (e.g., a parameterized optimization problem with a twice continuously differentiable objective function), even though – as we are well aware – none of our statements fundamentally depends on the differentiable structure. All of our results can be expected to also hold, if the problem is suitably discretized; nevertheless we prefer to adopt a differentiable approach for ease of exposition, since then the tools of differentiable geometry can be applied seamlessly. We decompose the problem of changing problem parameters to achieve monotone comparative statics (MCS) into two parts. *First*, the decision maker needs to solve a *local* MCS problem by finding for each point t in the parameter space $\mathcal{T}$ a direction vector $v(t)$ (i.e., an element of the tangent space at t) that would increase the unknown optimal action $x(t)$ if parameters were to be locally changed from t in the direction of $v(t)$. If the location of $x(t)$ is not known precisely, as is generally the case, then the direction $v(t)$ must be such that it induces local monotonicity with respect to all points in a subset $\mathcal{R}(t)$ of the action space $\mathcal{X}$ which is known to contain the optimal

[3]Even though not explicitly developed, our methods apply equally to equilibrium problems, by replacing the first-order necessary optimality conditions of the optimization problem with the equations specifying equilibria.

action $x(t)$, given the possibility that any point in $\mathcal{R}(t)$ might turn out to be optimal. A full solution to the first problem consists in a vector field $v(t)$ defined for all values of t in the parameter space. *Second*, given the vector field $v(t)$ the decision maker needs to find a reparameterization solving the *global* MCS problem. We show that this can always be achieved locally through "rectification" of the vector field. Under a few additional conditions rectification can also be achieved globally, leading to the desired global MCS reparameterization of the decision problem.

The chapter proceeds as follows. In Section 2.2 we introduce the problem of obtaining monotone comparative statics (i.e., monotone dependence of solutions on parameters) for parameterized optimization problems, both from a local and a global viewpoint. When considered locally, obtaining monotone comparative statics corresponds to finding directions in the parameter space in which solutions to the optimization problem are nondecreasing in parameters. A solution to the local monotone comparative statics problem is provided in Section 2.3 using a vector field method. Subsequently, in Section 2.4, we take a more global perspective: having obtained a vector field of monotone comparative statics directions defined at each point of the parameter space, we demonstrate that it is possible to (at least locally) change the parameters of the optimization problem, i.e., to *reparameterize* it, such that monotone comparative statics of the solutions of the reparameterized problem obtain. We show that the reparameterization can be global if a hyperplane can be found that is transverse to a vector field that solves the local MCS problem at each point of the parameter space. To illustrate our results we then discuss a number of applications in Section 2.5 before concluding with a discussion and directions for further research in Section 2.6.

2.2 Problem Formulation

We consider a decision maker who, given a parameter value $t \in \mathcal{T}$, aims to select an element $x(t)$ of an action space $\mathcal{X}$ so as to maximize her objective function $f : \mathcal{X} \times \mathcal{T} \to \mathbb{R}$. She thus tries to solve the *parameterized optimization problem*

$$x(t) \in \arg\sup_{x \in \mathcal{X}} f(x,t), \tag{1}$$

where $\mathcal{X}$ is an open subset[4] of $\mathbb{R}^n$ and $\mathcal{T}$ is an open subset of $\mathbb{R}^m$. Provided that a solution to (1) always exists,[5] the decision maker is interested in the comparative statics of the maximizer $x(t)$ as $t \in \mathcal{T}$ varies. More generally, we assume that the decision maker is concerned with the behavior of the composition $\varphi \circ x(t) = \varphi(x(t))$, where $\varphi : \mathcal{X} \to \mathbb{R}^d$ (with $1 \leq d \leq n$) is an *evaluation function* that the decision maker uses to assess any solution $x(t)$ that satisfies (1). For instance, if the decision maker is only interested in the comparative statics of the first component of the maximizer $x(t) = (x_1, \ldots, x_n)(t)$, she can choose $\varphi(x) = x_1$. From Milgrom and Shannon's (1994) Monotonicity Theorem, we know that if $\varphi(x) = x$ and $\mathcal{X}$ is a lattice, then $\varphi \circ x(t) = x(t)$ is increasing in t if and only if f is quasi-supermodular[6] in x and satisfies the single-crossing property[7] in (x,t). Conversely, for any objective

[4] If $\mathcal{X}$ lies in a lower-dimensional submanifold of $\mathbb{R}^n$, the analysis can still be applied, but differential calculus should be understood on this submanifold, and openness should be understood relative to the submanifold, cf. Section 2.3.5.

[5] If for any parameter $t \in \mathcal{T}$ the function $f(\cdot, t)$ is continuous and $\mathcal{X}$ is bounded, a solution to the parameterized optimization problem (1) exists in the closure of $\mathcal{X}$ by Weierstrass' Theorem (Bertsekas, 1995, p. 540).

[6] A real-valued function f defined on a lattice $\mathcal{X} \subset \mathbb{R}^n$ is quasi-supermodular if $f(x) \geq (>) f(x \wedge y)$ implies $f(x \vee y) \geq (>) f(y)$, for all x, y in $\mathcal{X}$, where $x \vee y = (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$ and $x \wedge y = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$. As its name suggests, quasi-supermodularity is a weaker condition than supermodularity.

[7] A real-valued function f defined on the product $\mathcal{X} \times \mathcal{T}$ of two partially ordered spaces has the single-crossing property if, whenever $x' > x$ and $t' > t$, $f(x', t) \geq (>) f(x, t)$ implies $f(x', t') \geq (>) f(x, t')$. The single-crossing property is a weaker condition than supermodularity in (x, t).

function f that does *not* satisfy these conditions together with a monotone evaluation function φ, the expression $\varphi \circ x(t)$ cannot be nondecreasing on $\mathcal{T}$. Nevertheless, despite this negative result using the standard theory, it may be possible for the decision maker to at least find a *path* γ_t in the parameter space, so that starting at a given $t \in \mathcal{T}$ the function $\varphi \circ x$ is nondecreasing along γ_t. Monotone comparative statics (MCS) may thus be obtained following certain directions in the parameter space. It is useful given any $t \in \mathcal{T}$ to state the decision maker's MCS problem (at t) in precisely these terms.

MCS PROBLEM (AT t). *Given a continuously differentiable evaluation function φ and a parameter value $t \in \mathcal{T}$, find a nonempty open interval $\mathcal{I}_t \subset \mathbb{R}$ with $0 \in \mathcal{I}_t$ and a path $\gamma_t : \mathcal{I}_t \to \mathcal{T}$, such that $\varphi \circ x(\gamma_t(\lambda))$ is nondecreasing*[8] *for all $\lambda \in \mathcal{I}_t$ and $\gamma_t(0) = t$.*

If a solution $(\mathcal{I}_t, \gamma_t)$ of the MCS problem at t is such that the path γ_t cannot be extended in $\mathcal{T}$, then we call the solution *maximal.* We can restrict our attention, without any loss of generality, to a maximal solution of the MCS problem at t. Monotone comparative statics relative to an evaluation function φ and a parameter starting value t obtain whenever the vector $\varphi \circ x$ is componentwise nondecreasing along an appropriate path γ_t in the parameter space $\mathcal{T}$. Along any such path, $x(\gamma_t(\lambda))$ solves (1) for all $\lambda \in \mathcal{I}_t$. A solution to the MCS problem for all $t \in \mathcal{T}$ results in a *global flow* $\theta(\lambda, t) = \gamma_t(\lambda)$, for which $\theta(0, t) = t$ and

$$\lambda \leq \mu \;\Rightarrow\; \varphi \circ x(\theta(\lambda, t)) \leq \varphi \circ x(\theta(\mu, t)), \tag{2}$$

[8] A vector $v(\lambda)$ is nondecreasing in λ if and only if each of its components is nondecreasing in λ.

for any $\lambda, \mu \in \mathcal{I}_t$. If it is possible to represent the global flow in the form

$$\theta(\lambda_t, \psi(t)) = t, \tag{3}$$

where λ_t is uniquely determined and $\psi : \mathcal{T} \to \mathcal{P}$ is a function that maps the parameter space to a fixed $(m-1)$-dimensional hypersurface $\mathcal{P}$ transverse[9] to the vector field induced by the MCS paths γ_t, then using the new parameters $s(t) = (s_1, \ldots, s_m)(t) = (\lambda_t, \pi(\psi(t)))$ guarantees monotone comparative statics of $\varphi \circ x(s)$ in $s_1 = \lambda$, at least locally, where π is a diffeomorphism from $\mathcal{P}$ onto a subset of $\mathbb{R}^{m-1}$ (details are provided in Section 2.4). Letting $\mathcal{F} = \bigcup_{t \in \mathcal{T}} \mathcal{I}_t \times \{t\}$, we formulate the global parameter-change problem accordingly.

Global MCS Reparameterization. *If the flow $\theta : \mathcal{F} \to \mathcal{T}$ solves the MCS problem everywhere in $\mathcal{T}$, find a new parameterization $s(t) = (s_1, \ldots, s_m)(t) = (\lambda_t, \pi(\psi(t)))$ such that (3) is satisfied for all $t \in \mathcal{T}$.*

A global MCS reparameterization provides the decision maker with new problem parameters $s = (s_1, ..., s_m)$ that guarantee monotonicity of $\varphi(\hat{x}(s))$ in the first component $s_1 \in \mathcal{I}$, where

$$\hat{x}(s) = \arg\max_{x \in \mathcal{X}} \hat{f}(x, s) \tag{4}$$

and $\hat{f}(x, s)$ corresponds to the objective function $f(x, t)$ after the parameter change.

[9]See Assumption 6 for the precise definition.

2.3 Solving the MCS Problem at t

Our goal is to find directions in which solutions of the parameterized optimization problem (1) (or functions thereof) are increasing. For this we introduce a "pseudo-gradient" $W(x,t)$ which mimics the gradient matrix $\nabla_t x(t) = W(x(t),t)$ corresponding to *all potential* solutions $x \in \mathcal{R}(t)$ with respect to the parameters t on a set $\mathcal{R}(t)$ that is known to contain the actual solution $x(t)$. We refer to $\mathcal{R}(t)$ as a "reduced feasible set," for it is a subset of the set of all feasible actions $\mathcal{X}$. The cardinality of $\mathcal{R}(t)$ is a measure of how much information the decision maker has about the location of the solution to (1) at t. If for a given $t \in \mathcal{T}$ all row-vectors of the pseudo-gradient lie in the same m-dimensional half-space for all points of the reduced feasible set (a subset of $\mathcal{X}$), then a direction $v(t) \in \mathbb{R}^m \setminus \{0\}$ exists in which monotone comparative statics obtain locally. Naturally, if such a direction can be found for all points t of the parameter space $\mathcal{T}$, then the resulting *vector field* $v : \mathcal{T} \to \mathbb{R}^m$ constitutes a *solution to the MCS problem on* $\mathcal{T}$. The flow induced by this vector field can then be used to obtain an MCS reparameterization of the optimization problem (1), which is discussed in Section 2.4. In what follows we first introduce a number of assumptions needed for the vector field method. We then provide techniques to implement the method, and relate the vector field method to classic supermodularity results. Finally, we provide important methods to deal with problems that contain equality and/or inequality constraints.

2.3.1 Preliminaries

In order to use standard tools from differential geometry, we require that f be sufficiently smooth.

Assumption 1 (Smoothness) *The objective function f is twice continuously differentiable in x and has continuous cross-derivatives with respect to each tuple (x_i, t_k), for all $1 \le i \le n$ and all $1 \le k \le m$.*

Let us denote by

$$H(x,t) = \nabla_{xx} f(x,t) = \left[\frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} \right]_{i,j=1}^{n}$$

the Hessian matrix of $f(\cdot, t)$ evaluated at (x,t), and by

$$K(x,t) = \nabla_{xt} f(x,t) = \left[\frac{\partial^2 f(x,t)}{\partial x_i \partial t_k} \right]_{i,k=1}^{n,m}$$

the matrix of cross-derivatives of f between decision-variable and parameter components, evaluated at (x,t). In order to bypass (at least for now) any difficulties arising from binding constraints at the optimal action, we will assume that the parameterized optimization problem (1) possesses a unique unconstrained optimum. Constrained optimization problems can often be restated equivalently so as to satisfy this assumption, and Section 2.3.5 is dedicated to this issue. We also emphasize that nothing in our method requires that the feasible set $\mathcal{X}$ be a lattice (cf. also footnote 19).

Assumption 2 (Existence and Uniqueness) *For each $t \in \mathcal{T}$, the parameterized optimization problem (1) has a unique solution $x(t)$.*

In general, the set of maximizers is guaranteed to be nonempty and in the interior of $\mathcal{X}$ if, in addition to being continuous, f is *coercive* relative to $\mathcal{X}$, in the sense that for any $t \in \mathcal{T}$ there exists a point $\check{x}(t) \in \mathcal{X}$ such that[10] $f(\check{x}(t), t) \ge \max f(\partial \mathcal{X}, t)$, cf. Bertsekas (1995, pp. 8, 540). If $f(\cdot, t)$ possesses multiple strict local extrema

[10] $\partial \mathcal{X}$ denotes the boundary of $\mathcal{X}$.

for some $t \in \mathcal{T}$, our results can be applied with respect to the comparative statics of each local maximum. Since $\mathcal{X}$ is open, Fermat's Lemma (Zorich, 2004, Vol. I, p. 215) implies that each strict local extremum $c(t)$ (and in particular the unique global maximum $x(t)$ guaranteed by Assumption 2) is a critical point of $f(\cdot, t)$, i.e., $\nabla_x f(c(t), t) = 0$. We denote by $\mathcal{C}(t)$ the *set of all critical points*[11] of $f(\cdot, t)$ in $\mathcal{X}$ at t,

$$\mathcal{C}(t) = \{x \in \mathcal{X} : \nabla_x f(x, t) = 0\}.$$

Thus, if the decision maker can determine $\mathcal{C}(t)$, she might be able to find the optimal action $x(t)$ as the solution of a *reduced* optimization problem, $x(t) = \arg\max_{c \in \mathcal{C}(t)} f(c, t)$, using the first-order necessary optimality conditions. Assumption 2 also implies that for any $t \in \mathcal{T}$ at the unique global optimum $x(t)$ the Hessian matrix of $f(\cdot, t)$ is negative semidefinite, i.e., $x(t)$ satisfies the *second-order necessary optimality condition* $x(t) \in \mathcal{D}(t)$, where

$$\mathcal{D}(t) = \{x \in \mathcal{X} : H(x, t) \leq 0\}.$$

This allows the decision maker to further reduce the optimization problem combining the first-order and second-order necessary optimality conditions and solve

$$x(t) = \arg\max_{x \in \mathcal{R}(t)} f(x, t), \tag{5}$$

where we refer to $\mathcal{R}(t) \subseteq \mathcal{X}$ as a *reduced feasible set*; in this case $\mathcal{R}(t) = (\mathcal{C} \cap \mathcal{D})(t)$. More generally, we refer to *any* subset $\mathcal{R}(t)$ of $\mathcal{X}$ which is guaranteed to contain

[11]Note that if $f(\cdot, t)$ has a critical point (i.e., $\mathcal{X} \cap \mathcal{C}(t) \neq \emptyset$) and is strictly concave on $\mathcal{X}$ for all $t \in \mathcal{T}$, then Assumption 2 is automatically satisfied, since $f(\cdot, t)$ is necessarily single-peaked on $\mathcal{X}$.

the solution $x(t)$ of (1) as an *admissible reduced feasible set.* Any element of $\mathcal{R}(t)$ is called a *reduced-feasible action.* If the decision maker can determine an admissible reduced feasible set $\mathcal{R}(t)$ (e.g., by using first- and second-order necessary optimality conditions), so that she is able to solve the (reduced) parameterized optimization problem (5) on $\mathcal{T}$, then the MCS problem always has a trivial solution, as will become clear below (cf. Theorem 1). Unfortunately, in many practical applications, a closed-form solution of (1) is not possible, or the objective function is not perfectly known by the decision maker (see Milgrom (1994) and Section 2.6). In that case, by constructing a reduced feasible set $\mathcal{R}(t) \subseteq \mathcal{X}$ that is guaranteed to contain the optimal action $x(t)$ (e.g., by using heuristics related to the special structure of the problem), the decision maker may still be able to solve the MCS problem without an explicit solution to the (equivalent) parameterized optimization problems (1) and (5). To obtain a solution to the MCS problem when the optimal action $x(t)$ can only be imperfectly localized in the set $\mathcal{R}(t) \subseteq \mathcal{X}$, we require that all critical points of $f(\cdot, t)$ in $\mathcal{R}(t)$ be nondegenerate (i.e., such that the Hessian matrix of $f(\cdot, t)$ is nonsingular there).

ASSUMPTION 3 (NONDEGENERACY) *For any $t \in \mathcal{T}$ the Hessian matrix $H(x,t)$ is nonsingular for all $x \in \mathcal{R}(t)$, for some reduced feasible set $\mathcal{R}(t) \subseteq \mathcal{X}$ which contains $x(t)$.*

This assumption guarantees that the inverse $H^{-1}(x,t)$ is well defined and continuous at any point $(x,t) \in \mathcal{R}(t) \times \mathcal{T}$. Hence the expression $-\left(H^{-1}K\right)(x,t)$, evaluated at a point (x,t) possibly *different* from the optimal $(x(t),t)$, is well-defined. Assumption 3 is automatically satisfied if the objective function is strictly concave.

LEMMA 1 *Under Assumptions 1–3 the unique optimal solution $x(t)$ of the parameterized optimization problem (1) is continuously differentiable on $\mathcal{T}$. The corresponding Jacobi matrix is given by*

$$\nabla_t x(t) = \left[\frac{\partial x_i(t)}{\partial t_k}\right]_{i,k=1}^{n,m} = -\left(H^{-1}K\right)(x(t),t), \tag{6}$$

for all $t \in \mathcal{T}$.

Proof. By Assumption 2, a unique interior solution $x(t)$ to the parameterized optimization problem (1) exists for all $t \in \mathcal{T}$, satisfying $H(x(t),t) \leq 0$ and

$$\nabla_x f(x(t),t) = 0. \tag{7}$$

By Assumption 1 we can differentiate (7) with respect to t (using the chain rule) and obtain

$$\nabla_t \left(\nabla_x f(x(t),t)\right) = (\nabla_{xx} f(x(t),t))(\nabla_t x(t)) + \nabla_{xt} f(x(t),t) = 0,$$

or equivalently

$$H(x(t),t)\nabla_t x(t) + K(x(t),t) = 0,$$

for all $t \in \mathcal{T}$. Since $H(x(t),t)$ is nonsingular by Assumption 3, we get expression (6) after left-multiplication with $H^{-1}(x(t),t)$ in the last equality. We now show that $\nabla_t x(t)$ is continuous. Since the maximizer $x(t)$ is unique and the objective function f continuous, we have that, as a consequence of Berge's (1963, p. 116) Maximum Theorem, the maximizer $x(t)$ is continuous in t (for it is upper-semicontinuous and single-valued). By virtue of Assumption 1 and nonsingularity of H, all entries of the matrix $\left(H^{-1}K\right)(x(t),t)$ are well defined and as a composition of continuous functions also continuous. Hence, the Jacobi matrix $\nabla_t x(t)$ on the left-hand side of equation (6) must also be continuous, which completes the proof. ■

In order to study the monotonicity of $\varphi \circ x(t)$, we require some smoothness as well as *functional independence* (Zorich, 2004, Vol. I, p. 508) of the evaluation function.

ASSUMPTION 4 (FUNCTIONAL INDEPENDENCE) *The evaluation function φ is continuously differentiable on $\mathcal{X}$, and its Jacobi matrix $\Phi(x) = \nabla_x \varphi(x) = \left[\frac{\partial \varphi_l(x)}{\partial x_i}\right]_{l,i=1}^{d,n}$ has (full) rank d for any (reduced-)feasible action x.*

Assumption 4 is not critical for our results and can (except for the smoothness portion) be relaxed. Functional independence guarantees that the MCS problem is locally never trivial, since no two of φ's components are collinear. If Assumptions 1–4 hold, then for any (reduced-)feasible tuple (x, t) we can define the $(d \times m)$-matrix

$$W(x,t) = -\Phi(x)H^{-1}(x,t)K(x,t), \tag{8}$$

which we term the *pseudo-gradient of the MCS problem at* (x, t). In analogy to Lemma 1, it is easy to show that the pseudo-gradient evaluated at any optimizing decision-parameter tuple $(x(t), t)$ describes the comparative statics of $\varphi \circ x(t)$ along paths parallel to the standard coordinates in the parameter space $\mathcal{T}$, i.e.,

$$\nabla_t \varphi(x(t)) = W(x(t), t). \tag{9}$$

We say that the pseudo-gradient of the MCS problem is *orientable* at (x, t), if the collection of all of its row vectors is a subset of a common half space of $\mathbb{R}^m$. If for a given $t \in \mathcal{T}$ the row vectors of $W(x, t)$ lie in a common half space of $\mathbb{R}^m$ for all x in a set $\mathcal{Y} \subseteq \mathcal{X}$, then we say that the *pseudo-gradient is orientable on $\mathcal{Y}$ at t.*

2.3.2 The Vector Field Method

To find a path γ_t that solves the MCS problem at $t \in \mathcal{T}$, our method requires tracking down some information about the direction of the gradient $\nabla_t(\varphi(x(t)))$, given that the decision maker only knows that $x(t)$ lies in some reduced feasible set $\mathcal{R}(t) \subseteq \mathcal{X}$. More specifically, we need to determine a direction forming acute angles with the pseudo-gradient of the MCS problem. Such a direction exists if the pseudo-gradient is orientable.

ASSUMPTION 5 (PSEUDO-GRADIENT ORIENTABILITY) *There exists a continuously differentiable vector field $v(t) \in \mathbb{R}^m \setminus \{0\}$ such that for each $t \in \mathcal{T}$,*

$$W(x,t)v(t) \geq 0, \tag{10}$$

for all $x \in \mathcal{R}(t)$, where $\mathcal{R}(t) \subseteq \mathcal{X}$ is an admissible reduced feasible set.

Under Assumption 5, $v(t)$ defines a vector field on $\mathcal{T}$ and a phase diagram with paths corresponding to the flow of this vector field.[12] The key result of this section is that $\varphi \circ x$ is nondecreasing along the paths.

THEOREM 1 *Under Assumptions 1-5, let $\mathcal{I} \subset \mathbb{R}$ be an open interval and $\gamma : \mathcal{I} \to \mathcal{T}$ a differentiable path such that*

$$\dot{\gamma}(\lambda) = v(\gamma(\lambda)),$$

for all $\lambda \in \mathcal{I}$. Then, $\varphi(x(\gamma(\lambda)))$ is nondecreasing for all $\lambda \in \mathcal{I}$.

[12] The trajectories of a vector field $v(t)$ exist and are unique on the whole domain $\mathcal{T}$ if the vector field is Lipschitz there (Khalil, 1992, pp. 74–77), and in particular when v is continuously differentiable on $\mathcal{T}$.

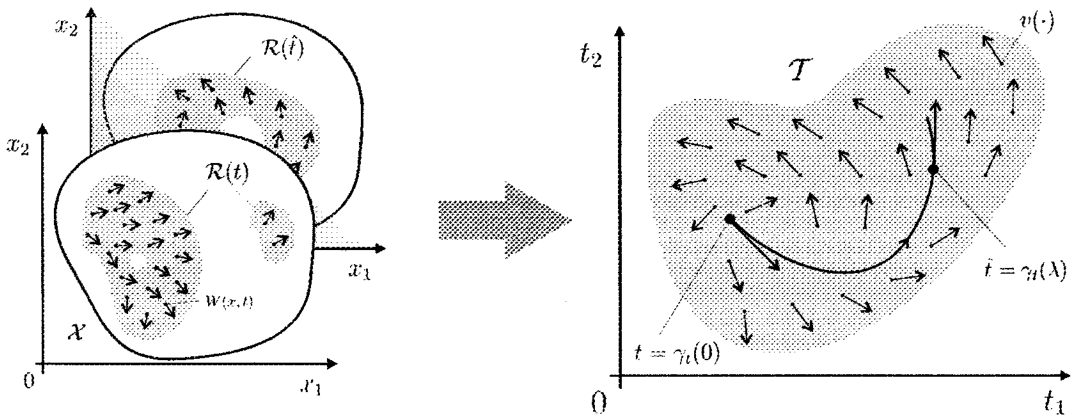

Figure 1: Solution to the MCS Problem at t for $\varphi(x) = x_1$.

Proof. Since γ is differentiable on $\mathcal{I}$, Lemma 1 implies (using Assumptions 1 and 2) that

$$\nabla_t x(\gamma(\lambda)) = -\left(H^{-1}K\right)(x(\gamma(\lambda)), \gamma(\lambda)).$$

Thus, using Assumption 4 and equation (9), the gradient of $\varphi(x(\gamma(\lambda)))$ with respect to λ is given by

$$\nabla_\lambda \varphi(x(\gamma(\lambda))) = W(x(\gamma(\lambda)), \gamma(\lambda))\dot{\gamma}(\lambda) = W(x(\gamma(\lambda)), \gamma(\lambda))v(\gamma(\lambda)).$$

By virtue of inequality (10) in Assumption 5, the latter expression is nonnegative.■

Under the assumptions of Theorem 1, we say that $\varphi(x(t))$ *is nondecreasing along the trajectories of* $v(t)$.

2.3.3 Implementation

Under Assumptions 1–5 the vector field method can be implemented using the following techniques. *First*, for any $t \in \mathcal{T}$ determine a reduced feasible set $\mathcal{R}(t) \subseteq \mathcal{X}$ that is guaranteed to contain the interior solution $x(t)$ of the parameterized optimization problem (1). *Second*, to satisfy Assumption 5, verify that the pseudo-gradient $W(x,t)$ is orientable on $\mathcal{R}(t)$. Note that for any given $t \in \mathcal{T}$ it may be possible to achieve orientability by premultiplying the evaluation function φ by some diagonal $(d \times d)$-matrix of the form $M(t) = \text{diag}(m_1, ..., m_d)(t)$ where $m_l(t) \in \{-1, 1\}$ for all $l \in \{1, \ldots, d\}$. Indeed if $\hat{\varphi}(x,t) = M(t)\varphi(x)$, then it is possible to find a matrix $M(t)$ such that the pseudo-gradient $\hat{W}(x,t)$ corresponding to the modified evaluation function $\hat{\varphi}$,

$$\hat{W}(x,t) = M(t)W(x,t) = -M(t)\Phi(x)H^{-1}(x,t)K(x,t),$$

is orientable at (x,t). In particular, if a matrix M can be found that is independent of t, it may be advantageous for the analysis of the problem if the decision maker uses the evaluation function $\hat{\varphi}$ instead of φ. *Third,* find a vector field v that satisfies (10). To accomplish this, a systematic, algorithmic procedure to determine a vector field $v(t)$ that is "maximally aligned" with the pseudo-gradient $W(x,t)$ consists in solving the maximin problem,[13]

$$v(t) \in \arg \max_{v \in \mathbb{R}^m : \|v\|=1} \left\{ \min_{x \in \mathcal{R}(t)} \left\{ \min_{1 \le l \le d} \langle W_l(x,t), v \rangle \right\} \right\}. \tag{11}$$

When the assumptions of the Min-Max Theorem (Kakutani, 1942) are satisfied, any solution $v(t)$ to problem (11) also satisfies

$$v(t) \in \arg \min_{x \in \mathcal{R}(t)} \left\{ \max_{v \in \mathbb{R}^m : \|v\|=1} \left\{ \min_{1 \le l \le d} \langle W_l(x,t), v \rangle \right\} \right\}. \tag{12}$$

[13] W_l is the l–th row vector of W, and $\langle \cdot, \cdot \rangle$ denotes the scalar product in the relevant Hilbert space.

Motivated by the minimax formulation (12), since W is orientable by assumption, one obtains the set of candidate vector fields

$$\mathcal{V} = \left\{ \frac{W_1(\hat{x}^1(t),t)}{\|W_d(\hat{x}^1(t),t)\|}, \dots, \frac{W_d(\hat{x}^d(t),t)}{\|W_d(\hat{x}^d(t),t)\|} \right\}\Bigg|_{t\in\mathcal{T}}, \tag{13}$$

where for any $l \in \{1,\dots,d\}$ we have set

$$\hat{x}^l(t) = \arg\min_{x\in\mathcal{R}(t)} \|W_l(x,t)\|.$$

One can now check whether some element of $\mathcal{V}$ is a suitable vector field on $\mathcal{T}$, or possibly a subset thereof.

The above three steps can be iterated to tighten the reduced feasible set. It can also be useful to only consider subsets of the parameter space $\mathcal{T}$. Note that if an exact solution to the maximization problem (5) is known for some $t \in \mathcal{T}$, then the MCS problem at t has a solution if any only if $W(x(t),t)$ is orientable at $(x(t),t)$.

EXAMPLE 1 Consider a firm that has the option to invest in a number x of geographically dispersed markets[14] (e.g., cities in the US) at an increasing convex cost $C(x) \geq 0$. For simplicity, the market price $p \in (0,1)$ for the firm's product is assumed to be the same in each market; it is announced nationally and is a parameter of the problem. Marginal costs for hamburgers are zero. The demands in the different markets are uncorrelated and the firm is risk averse with constant absolute risk aversion ρ. By investing in x markets the firm also reaps an increasing concave side benefit $B(x) \geq 0$ (e.g., through real-estate transactions).[15] With a probability $q \in (0, 1-p)$ that is

[14]We allow x to take non-integer values.

[15]By imposing the Inada conditions $C'(0) < \infty$ and $B'(0) = \infty$ one can easily guarantee that the

at most directly proportional to the quantity sold in each market, the firm incurs a unit loss in any market due to a liability claim. The firm's expected payoffs are approximately[16]

$$\begin{aligned}\Pi(x,q,p) &= p(1-p)x + B(x) - E\tilde{L} - \frac{\rho}{2}\mathrm{var}(\tilde{L}) - C(x)\\ &= \left(p(1-p) - q - \frac{\rho q(1-q)}{2}\right)x + B(x) - C(x),\end{aligned}$$

where the random variable $\tilde{L} \in \{0, 1, \ldots, x\}$ represents the firm's total losses. Maximizing profits we thus obtain $x^*(q,p) = \xi(p(1-p) - q - \rho q(1-q)/2)$, where ξ is the (increasing) inverse of $C' - B'$. Hence, the maximizer $x(t)$ is increasing in the parameter $t \in \mathcal{T} = \{(q,p) \in \mathbb{R}^2_{++} : p + q < 1\}$ if and only if $p(1-p) - q - \rho q(1-q)/2$ is increasing. The corresponding pseudo-gradient at the optimum is

$$W(x(t)) = [-1 + \rho q - \rho/2, 1 - 2p]\, \xi'(p(1-p) - q - \rho q(1-q)/2),$$

so that with $v(t) = (-p, q\rho/2)$ we obtain

$$\langle W(x(t)), v(t)\rangle = \frac{\rho\xi'(t)}{2}\left(\frac{2p}{\rho} + p + q - 4pq\right) \geq \frac{(4\rho - 1)\xi'(t)}{16} > 0,$$

for all $t \in \mathcal{T}$ and $\rho > 1/4$. The simple elliptic vector field v thus solves the MCS problem on $\mathcal{T}$. In Example 3 we show how to obtain a global MCS reparameterization of the problem based on the vector field v, cf. Figure 2. Let us remark that clearly in this example $\mathcal{R}(t) = \{x(t)\}$, if ξ^{-1} is known precisely (for any given B and C). However,

optimum is interior, i.e., $x(t) > 0$, so that we can without loss of generality set $\mathcal{X} = (0, \infty)$.

[16]The dependence of Π on the parameter ρ is not explicitly noted. In fact, in this problem the comparative statics with respect to ρ are obvious. We can thus use ρ itself in solving the MCS problem and finding an appropriate simple reparameterization, which illustrates an interesting "partial reparameterization" variant of our technique.

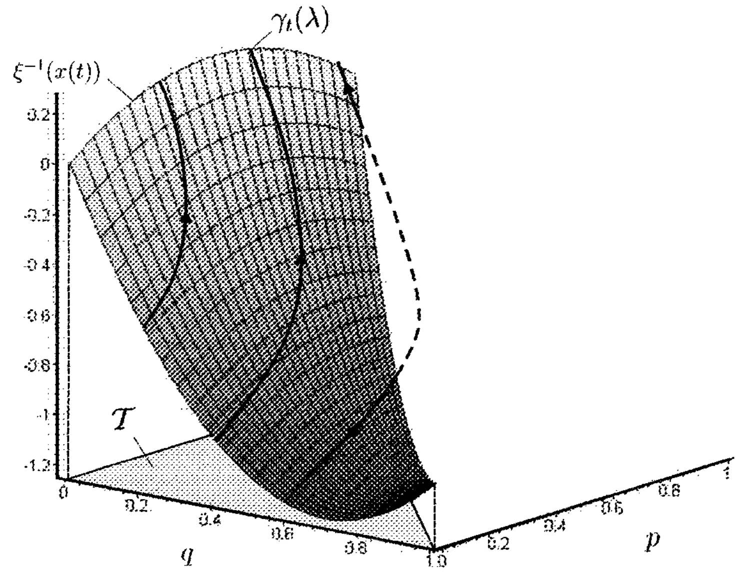

Figure 2: Global MCS Reparameterization in Examples 2 and 3.

our conclusions can be obtained without further specifying the firm's payoffs. Also note that it would have been possible to trivially replace $p(1-p)-q-\rho(1-q)q/2$ by a scalar parameter λ, which would somewhat decrease the resolution of insight (not a one-to-one mapping from the parameter space) for the decision maker somewhat, yet clearly provide trivial but precise monotone comparative statics. □

2.3.4 Relation to Classic Supermodularity Results

We now derive a well-known supermodularity result as a particular case of Theorem 1. Under Assumption 1, we recall that $f(x,t)$ is *supermodular in* x if $\frac{\partial^2 f(x,t)}{\partial x_i \partial x_j}$ is nonnegative for all $(x,t) \in \mathcal{X} \times \mathcal{T}$ and $1 \leq i \neq j \leq n$. We call the function $f(x,t)$ *supermodular in* (x,t) if in addition $\frac{\partial^2 f}{\partial x_i \partial t_k}(x,t)$ is nonnegative for all $(x,t) \in \mathcal{X} \times \mathcal{T}$ and $1 \leq i \leq n, 1 \leq k \leq m$.

COROLLARY 1 *Suppose that Assumptions 1-3 hold and that f is supermodular in (x,t). Then $x(t)$ is nondecreasing in t.*

Proof. We show that $x(t)$ is nondecreasing in each component of $t = (t_1, \ldots, t_m)$. Supermodularity in (x,t) implies that all components of $K(x,t)$ are nonnegative on $\mathcal{X} \times \mathcal{T}$. It also implies that $H_{ij}(x,t) \geq 0$ for all $i,j \in \{1,\ldots,m\}$ with $i \neq j$ on $\mathcal{X} \times \mathcal{T}$. Without loss of generality we can restrict our attention to the reduced feasible set $\mathcal{R}(t) = (\mathcal{C} \cap \mathcal{D})(t)$. Hence $H_{ii}(x(t),t) \leq 0$ for all $i \in \{1,\ldots,n\}$, for the Hessian matrix is negative definite at the optimum. Since $H_{ij}^{-1} = (-1)^{i+j}\det(H^{ji})/\det(H)$, it is a simple linear algebra exercise to verify that $H_{ij}^{-1} \leq 0$ on $\mathcal{X} \times \mathcal{T}$ for all i,j in $\{1,\ldots,n\}$.[17] For any vector $v(t) > 0$ all entries of $-H^{-1}(x,t)K(x,t)v(t)$ are therefore nonnegative. As a result, Assumption 5 is satisfied for $\varphi(x) = x$ and $v(t) \equiv e_k$ where e_k is the k-th unit vector in the canonical basis of $\mathbb{R}^m$. An application of Theorem 1 with $\varphi(x) = x$ concludes the proof. ■

An important case that is not currently dealt with in the monotone comparative statics literature is when f is supermodular in x but does not have the single-crossing property in (x,t).[18] In that context, Assumption 5 can be simplified as follows.

ASSUMPTION 5' *For each $t \in \mathcal{T}$, there exists a vector $v(t) \in \mathbb{R}^m$ such that $K(x,t)v(t)$ is nonnegative for all $x \in \mathcal{R}(t)$, where $\mathcal{R}(t) \subseteq \mathcal{X}$ is an admissible reduced feasible set.*

COROLLARY 2 *Suppose that Assumptions 1-3 and 5' hold and that f is supermodular in x. Then, $x(t)$ is nondecreasing along the trajectories of v.*

[17] The adjoint matrix H^{ij} is obtained by removing the i-th row and the j-th column from H.
[18] In particular, f is not supermodular in (x,t).

Proof. We modify the proof of Corollary 1. Supermodularity in x ensures that $H_{ij}^{-1} \leq 0$ on $\mathcal{X} \times \mathcal{T}$ for all $i, j \in \{1, \ldots, n\}$. This together with Assumption 5' implies that all entries of $-H^{-1}(x,t)K(x,t)v(t)$ are nonnegative. A direct application of Theorem 1 with $\varphi(x) = x$ concludes the proof. ■

Corollary 2 applies to situations in which there are complementarities between the different decision variables, but not between decision variables and parameters. Since the maximizer under the original parameterization can be nonmonotonic, it is clear that Assumption 5' relaxes the tight single-crossing requirement put forward in Milgrom in Shannon (1994) for the price of an MCS reparameterization of the problem.

2.3.5 Constrained Optimization Problems

Assumption 2 requires that the optimizer be in the *interior* of the feasible set $\mathcal{X}$. This assumption can be relaxed in different ways, *either* by reducing the dimensionality of the decision space (using a substitution approach for equality constraints) *or* by augmenting the dimensionality of the decision space (using a Lagrange-multiplier approach for equality and/or inequality constraints).

Substitution Approach. Any equality constraints that are part of the definition of the feasible set $\mathcal{X}$ define in fact a lower-dimensional set $\mathcal{X}'$ that forms a submanifold of $\mathcal{X}$ (with or without boundary). If the equality constraints can be solved globally for a number of decision variables, the parameterized optimization problem can be viewed as unconstrained on $\mathcal{X}'$ after backsubstitution of these variables. More specifically, if $\mathcal{X}'$ is diffeomorphic to an open subset of $\mathbb{R}^{\hat{n}}$ with $\hat{n} < n$, the problem can be seen as unconstrained on an open subset of $\mathbb{R}^{\hat{n}}$. To render our discussion precise, consider the problem[19]

$$\max_{x \in \mathcal{X}(t)} f(x,t), \tag{14}$$

with

$$\mathcal{X}(t) = \{x \in \mathcal{Y} : g(x,t) = 0\},$$

where $\mathcal{Y}$ is an open subset of $\mathbb{R}^n$, t belongs to an open set $\mathcal{T}$ of $\mathbb{R}^m$, and g takes values in $\mathbb{R}^k$ (for some $1 \leq k < n$) and is twice continuously differentiable. Suppose that the level set $g(x,t) = 0$ can be expressed explicitly as $(x_{n-k+1}, \ldots, x_n)(t) = \tilde{g}(x_1, \ldots, x_{n-k}, t)$, for $(x_1, \ldots, x_{n-k}) \in \mathcal{X}'$, where $\mathcal{X}'$ is the projection of $\mathcal{Y}$ on the plane $\{(x_1, \ldots, x_{n-k}, 0, \ldots, 0) : (x_1, \ldots, x_{n-k}) \in \mathbb{R}^{n-k}\} \subset \mathbb{R}^n$. The problem is then reduced to the $(n-k)$-variable unconstrained problem on $\mathcal{X}'$

$$\max_{(x_1,\ldots,x_{n-k}) \in \mathcal{X}'} \tilde{f}(x_1, \ldots, x_{n-k}, t) = \max_{(x_1,\ldots,x_{n-k}) \in \mathcal{X}'} f(x_1, \ldots, x_{n-k}, \tilde{g}(x_1, \ldots, x_{n-k}, t), t).$$

The application discussed in Section 2.5.3 provides a simple example of this transformation.

Lagrange-Multiplier Approach. Equality constraints can also be approached with Lagrange multipliers. In the previous example, a necessary condition (Bertsekas, 1995, p. 255) for optimality is the existence of a k-dimensional vector ν such that, at the optimum,

$$\nabla_x f(x,t) + \nu^T \nabla_x g(x,t) = 0.$$

[19]Note that in this formulation it is possible to have the feasible set depend on parameters. In contrast to standard MCS results obtained on lattices, we do not assume at the outset that $\mathcal{X}(t)$ is monotone in t with respect to the Veinott set order (Milgrom and Shannon, 1994). We are grateful to Pete Veinott for pointing out that his set order (originally termed "lower than" relation) was first introduced by him in a 1965 unpublished paper.

Together with the k equations $g(x,t) = 0$, this determines the system of $n+k$ equations in $n+k+m$ variables x, t, and ν

$$G(x,t,\nu) = \begin{bmatrix} \nabla_x f(x,t) + \nu^T \nabla_x g(x,t) \\ g(x,t) \end{bmatrix} = 0.$$

The implicit function theorem implies that, if $G_{x\nu}$ is invertible, then locally

$$\nabla_t(x,\nu)(t) = - \left[G_{x\nu}^{-1}\, G_t\right]\ (x,\nu,t).$$

Even though the position of the optimal x and ν in $\mathcal{X} \times \mathbb{R}^k$ is unknown, it might be possible to find directions in the parameter space, such that $x(t)$ is nondecreasing in these directions. The following example illustrates this Lagrange-multiplier approach with equality constraints.

EXAMPLE 2 Consider an economy with two goods $(x,y) \in \mathbb{R}_+^2$, with the production frontier $\{(x,y) \in \mathbb{R}_+^2 : g(x,y) = x^2 + y^2 - 1 = 0\}$ and a representative agent with utility $f(x,y) = u(x) + tv(y)$, where u, v are twice continuously differentiable, increasing and concave. We would like to determine the monotonicity properties of the optimizer $(x,y)(t)$ with respect to the parameter t. The constraint set is clearly not a lattice, hence classic supermodularity results do not apply directly.[20] Using the Lagrange-multiplier approach we have

$$G(x,y,\nu,t) = \begin{bmatrix} u'(x) + 2\nu x \\ tv'(y) + 2\nu y \\ x^2 + y^2 - 1 \end{bmatrix},$$

[20]However, classic results could be applied in conjunction with the substitution approach described earlier.

which implies that

$$G_{x,y,\nu}(x,y,\nu,t) = \begin{bmatrix} u''(x)+2\nu & 0 & 2x \\ 0 & tv''(y)+2\nu & 2y \\ 2x & 2y & 0 \end{bmatrix}$$

and that

$$G_t(x,y,\nu,t) = \begin{bmatrix} 0 \\ v'(y) \\ 0 \end{bmatrix}.$$

An application of the implicit function theorem then yields

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ \nu \end{bmatrix} = \frac{v'(y)}{D}\begin{bmatrix} -4xy \\ 4x^2 \\ 2y(u''(x)+2\nu) \end{bmatrix},$$

where $D = -4y^2(u''(x)+2\nu) - 4x^2(tv''(y)+2\nu)$. In this problem, the condition $\nabla f + \nu\nabla g = 0$ implies that ν is negative, since the gradients of f and g both belong to the positive orthant of $\mathbb{R}^2$. This, along with the concavity of u and v, implies that D is positive and that $\nabla_t x < 0 < \nabla_t y$, i.e., monotone comparative statics obtain. □

Inequality constraints can be approached in a similar fashion. Consider again problem (14), this time with

$$\mathcal{X}(t) = \{x \in \mathcal{Y} : g(x,t) = 0, h(x,t) \leq 0\},$$

where $\mathcal{Y}$ is an open subset of $\mathbb{R}^n$, t belongs to an open set $\mathcal{T}$ of $\mathbb{R}^m$, g and h take values in respectively $\mathbb{R}^k$ and $\mathbb{R}^r$ (with $k < n$) and are both twice continuously differentiable.

The Kuhn-Tucker necessary optimality conditions (Bertsekas, 1995, p. 284) imply the existence of adjoint variables ν and μ in respectively $\mathbb{R}^k$ and $\mathbb{R}^r_+$, such that

$$\nabla_x f + \langle \nu, \nabla_x g \rangle + \langle \mu, \nabla_x h \rangle = 0,$$

and

$$\hat{\mu}_i(x, \mu, t) = \mu_i(t) h_i(x, t) = 0$$

for all $i \in \{1, \ldots, r\}$. Letting

$$G(x, \nu, \mu, t) = \begin{bmatrix} \nabla_x f + \nu^T \nabla_x g + \mu^T \nabla_x h \\ g(x, t) \\ \hat{\mu}(x, \mu, t) \end{bmatrix},$$

a necessary optimality condition is that $G(x, \nu, \mu, t) = 0$. This defines a system of $n+k+r$ equations in $n+k+r+m$ variables, so that we can apply (if the relevant matrix is invertible) the implicit function theorem to compute $\nabla_t(x, \nu, \mu)$, and proceed as in the equality case. In some problems, it is possible to know in advance which inequality constraints are binding at the optimum. In this simple case, non-binding inequalities are ignored, while binding ones are treated as equality constraints. This approach is illustrated in the applications of Sections 2.5.2 and 2.5.3.

2.4 Finding a Global MCS Reparameterization

Given a smooth solution $v : \mathcal{T} \to \mathbb{R}^m$ to the MCS Problem, it is interesting in practice to find an MCS reparameterization of the optimization problem (1). The idea is to start with the flow $\theta : \mathcal{F} \to \mathcal{T}$ induced by the vector field v and note that this flow is smooth and unique on what we refer to as the maximum "flow domain" $\mathcal{F} \subset \mathbb{R} \times \mathcal{T}$, beyond which the integral curves of the vector field cannot be extended. By taking a plane that is transverse (i.e., never collinear) to these integral curves, it is possible to construct new parameter coordinates under which monotone comparative statics obtain, at least locally.

2.4.1 Global Flows

As a consequence of the standard theory of ordinary differential equations (ODEs; Arnold, 1973), if the solution v is smooth, then integral curves to the vector field exist, are unique, and induce a smooth local flow θ.[21] To define the concept of a local flow, let us first introduce a *flow domain* $\mathcal{F} \subset \mathbb{R} \times \mathcal{T}$ with the property that for any $t \in \mathcal{T}$, the set

$$\mathcal{F}^{(t)} = \{\lambda \in \mathbb{R} : (\lambda, t) \in \mathcal{F}\} \subset \mathbb{R}$$

is an open interval containing zero. A *local flow* on $\mathcal{T}$ is a continuous map $\theta : \mathcal{F} \to \mathcal{T}$ that satisfies the two group laws:

$$\theta(0, t) = t, \tag{15}$$

[21] Existence and uniqueness of integral curves is also obtained when the vector field v merely satisfies a Lipschitz condition (cf. footnote 12).

for all $t \in \mathcal{T}$, and

$$\theta(\lambda, \theta(\mu, t)) = \theta(\lambda + \mu, t), \tag{16}$$

for all $\lambda \in \mathcal{F}^{(t)}$ and $\mu \in \mathcal{F}^{(\theta(\lambda,t))}$ such that $\lambda + \mu \in \mathcal{F}^{(t)}$. In addition, for a given local flow θ, we define

$$\theta_\lambda(t) = \theta^{(t)}(\lambda) = \theta(\lambda, t)$$

whenever $(\lambda, t) \in \mathcal{F}$. A local flow (sometimes also referred to as a "one-parameter group action") relates for any $t \in \mathcal{T}$ the vector field $v(t)$ to its orbits $\theta^{(t)}(\lambda)$. If the flow domain is such that the map θ is surjective (i.e., $\theta(\mathcal{F}) = \mathcal{T}$), then we call θ a *global flow.*

THEOREM 2 (ODE EXISTENCE, UNIQUENESS, AND SMOOTHNESS) *Let $v : \mathcal{T} \to \mathbb{R}^m$ be a smooth vector field. Consider the initial value problem*

$$\nabla_\lambda \gamma_t(\lambda) = v(\gamma_t(\lambda)), \quad \gamma_t(\lambda_0) = t. \tag{17}$$

(i) EXISTENCE: *For any $\lambda_0 \in \mathbb{R}$ there exist a nonempty open interval $\mathcal{I}$ which contains λ_0 and an open set $\mathcal{U} \subset \mathcal{T}$ such that for any $t \in \mathcal{U}$ there is a smooth integral curve $\gamma_t : \mathcal{I} \to \mathcal{T}$ which solves (17) for all $\lambda \in \mathcal{I}$.*

(ii) UNIQUENESS: *Any two smooth solutions to (17) agree on their common domain.*

(iii) SMOOTHNESS: *Let $\mathcal{F} = \mathcal{I} \times \mathcal{U}$ as in (i). If we define the local flow $\theta : \mathcal{F} \to \mathcal{T}$ with $\theta(\lambda, t) = \gamma_t(\lambda)$, then θ is smooth.*

Proof. See e.g., Lee (2003, pp. 452–459). ■

In the following we use global flows corresponding to the solution vector field v to the MCS problem on $\mathcal{T}$ (or a subset thereof) in order to find a global MCS reparameterization of problem (1) as previously indicated in Section 2.2

2.4.2 Coordinate Change in Parameter Space

By construction, if v is a solution to the MCS problem on $\mathcal{T}$, it is nonsingular everywhere. The lack of singular points allows us to give a canonical, local representation of v using a change of coordinates in $\mathcal{T}$. Consider a point t_0 of $\mathcal{T}$. If v is smooth, there exists an open interval $\mathcal{I} \subset \mathbb{R}$ containing the origin and an open subset $\mathcal{U}$ of $\mathcal{T}$ satisfying the conclusions of Theorem 2 (with $\lambda_0 = 0$). Moreover, since $v(t_0)$ is nonsingular, it uniquely determines an orthogonal hyperplane $\mathcal{H} \subset \mathbb{R}^m$ containing t_0. Let

$$\mathcal{P} = \mathcal{U} \cap \mathcal{H} \subset \mathbb{R}^m.$$

Since $\mathcal{H}$ is diffeomorphic to $\mathbb{R}^{m-1}$, $\mathcal{P}$ can also be seen as an open subset of $\mathbb{R}^{m-1}$. To avoid confusion, let $\Pi = \pi(\mathcal{P})$ denote the image of $\mathcal{P}$ under the diffeomorphism $\pi : \mathcal{H} \to \mathbb{R}^{m-1}$. Theorem 2 implies the existence of a smooth flow $\theta(\lambda, t)$, which we restrict to the domain $\mathcal{I} \times \mathcal{P}$. The flow can be reparameterized, with a slight abuse of notation, by $\theta(\lambda, \pi)$ on the product $\mathcal{S} = \mathcal{I} \times \Pi \subset \mathbb{R}^m$. Moreover, the corresponding range $\bar{\mathcal{T}} = \theta(\mathcal{S})$ is an open subset of $\mathcal{T}$ by Theorem 2. Last, θ is one-to-one and smooth from $\mathcal{S}$ to $\bar{\mathcal{T}}$, also by Theorem 2. We have therefore defined a local change of coordinates around t_0: any t in the open neighborhood $\bar{\mathcal{T}}$ of t_0 can be uniquely expressed by a tuple $(\lambda_t, \pi(\psi(t))) \in \mathcal{S}$. The $(m-1)$-dimensional component $\psi(t)$ of t is the intersection of the trajectory going through t with the hyperplane $\mathcal{H}$. Moreover, since $t = \gamma_{\psi(t)}(\lambda_t)$, Theorem 1 implies that $\varphi \circ x$ is nondecreasing along the trajectories of the flow θ as λ increases. We have thus proved the following result.

THEOREM 3 (LOCAL MCS REPARAMETERIZATION) *If v solves the MCS problem on $\mathcal{T}$, there exists a local MCS reparameterization of the form*

$$s = (s_1(t), s_2(t), ..., s_m(t)) = (\lambda_t, \pi(\psi(t))), \tag{18}$$

around any $t_0 \in \mathcal{T}$, such that $\varphi(x(s))$ is nondecreasing in s_1.

In practice, this result can often be applied globally as the following example illustrates.

EXAMPLE 3 In Example 1 we obtained a vector field of the form $v(t) = (-t_2, \alpha t_1)$ (for some $\alpha > 0$) as the solution to an MCS problem on some $\mathcal{T} \subset \mathbb{R}^2_+ \setminus \{0\}$. Since $v_2(t) = \alpha t_1 \neq 0$ on $\mathcal{T}$, we can choose the plane $\mathcal{P} = \{t \in \mathcal{T} : t = (t_1, 0)\}$, which is transverse to the vector field. The (global) flow of the vector field $v(t)$ is

$$\theta_\lambda(t_1, t_2) = \left(t_1 \cos\sqrt{\alpha}\lambda - \frac{t_2 \sin\sqrt{\alpha}\lambda}{\sqrt{\alpha}}, t_1\sqrt{\alpha}\sin\sqrt{\alpha}\lambda + t_2\cos\sqrt{\alpha}\lambda\right).$$

Thus, for any $(s, 0) \in \mathcal{P}$, we obtain

$$\theta_\lambda(s, 0) = (s\cos\sqrt{\alpha}\lambda, s\sqrt{\alpha}\sin\sqrt{\alpha}\lambda),$$

for $\lambda \in (0, \pi/(2\sqrt{\alpha}))$ and $s > 0$. Hence, on any contractible[22] compact subset $\bar{\mathcal{T}}$ of $\mathcal{T}$, we obtain the global reparameterization $t \mapsto (\lambda, s)$ with $\lambda = \frac{1}{\sqrt{\alpha}} \arctan \frac{t_2}{\sqrt{\alpha}t_1}$ and $s = \sqrt{t_1^2 + (t_2^2/\alpha)}$, cf. Figure 2. In the context of Example 1, the key insight for the decision maker from the MCS reparameterization is that the optimal number of markets to invest in varies monotonically in the ratio p/q, i.e., the product price in

[22]An m-dimensional open set with nonempty interior is contractible if it is homotopy equivalent (i.e., it can be deformed via a continuous transformation) to an m-dimensional open ball. Intuitively, contractible sets have no "holes."

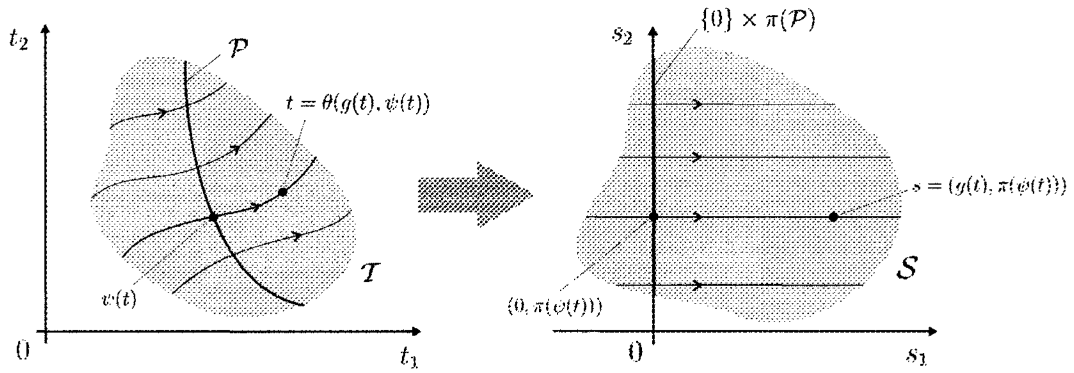

Figure 3: Global MCS Reparameterization.

relation to the risk of liability lawsuits. □

We now formally generalize the reparameterization technique used in the local case by providing a general condition under which Theorem 3 holds globally.

ASSUMPTION 6 (TRANSVERSE HYPERSURFACE) *There exist subsets* $\bar{\mathcal{T}}$ *and* $\mathcal{P}$ *of* $\mathcal{T}$, *with* $\bar{\mathcal{T}}$ *open and* $\mathcal{P} \subset \bar{\mathcal{T}}$, *such that:*

(i) there exists a convex, open subset Π *of* $\mathbb{R}^{m-1}$ *and a diffeomorphism* π *mapping* $\mathcal{P}$ *onto* Π;

(ii) for each $t \in \bar{\mathcal{T}}$, $\gamma_t(\mathcal{F}^{(t)}) \cap \mathcal{P}$ *is a singleton* $\{\psi(t)\}$.

This last assumption ensures that trajectories of an MCS vector field v lead to a foliation of the subset $\bar{\mathcal{T}}$ of $\mathcal{T}$. In other words, the existence of a set $\mathcal{P}$ of points, each element of which can be associated with exactly one trajectory, allows projection of the set $\bar{\mathcal{T}}$ onto $\mathcal{P}$ and – via the length (from t to $\psi(t)$) of the trajectory (which could pass outside $\bar{\mathcal{T}}$) – obtain a bijection between $\bar{\mathcal{T}}$ and a set $\mathcal{S} \subset \mathbb{R}^m$. This bijection corresponds to the desired global MCS reparameterization containing the length of the MCS trajectories as one new parameter in which the solution to the reparameterized problem (4) (when evaluated with φ) varies monotonically.

THEOREM 4 (GLOBAL MCS REPARAMETERIZATION) *If the vector field v is a solution to the MCS problem on $\bar{\mathcal{T}}$ and Assumption 6 holds, then there exists a global MCS reparameterization of the form*

$$s = (s_1(t), s_2(t), ..., s_m(t)) = (\lambda_t, \pi(\psi(t))), \tag{19}$$

such that $\varphi(x(s))$ is nondecreasing in s_1.

Proof. Each element t of $\bar{\mathcal{T}}$ uniquely determines an element $\psi(t)$ of $\mathcal{P}$ and a real λ_t such that $\gamma_{\psi(t)}(\lambda_t) = t$. The set $\mathcal{S} = \bigcup_{\psi \in \mathcal{P}} \mathcal{F}^{(\psi)} \times \{\pi(\psi)\}$ is an open subset of $\mathbb{R}^m$. Moreover, the mapping $t \mapsto s = (\lambda_t, \pi(\psi(t))) \in \mathcal{S}$ is one-to-one from $\bar{\mathcal{T}}$ to $\mathcal{S}$, and smooth by Theorem 2. Last, since $t(s) = (\lambda, \psi)$ follows the trajectories of v as λ increases, Theorem 1 implies that $\varphi(x(s)) = \varphi(x(\lambda, \pi))$ is nondecreasing in $s_1 = \lambda$.■

When v has a potential[23] $u : \mathcal{T} \to \mathbb{R}$, a good candidate for $\mathcal{P}$ in Assumption 6 is any iso-potential that crosses all trajectories.[24] Although this need not always be the case, it is likely that iso-potentials will be diffeomorphic to an open subset of $\mathbb{R}^{m-1}$, and to a convex subset if one chooses $\bar{\mathcal{T}}$ carefully. The following classic theorem (essentially Poincaré's Lemma) gives a necessary and sufficient condition for the existence of a potential, which can be tested on any vector field v satisfying Assumption 5.

[23] That is, v is the gradient of u.

[24] In that case, trajectories cross $\mathcal{P}$ only once, because the potential increases along them.

THEOREM 5 (EXISTENCE OF A POTENTIAL) *Let $\bar{\mathcal{T}} \subset \mathcal{T}$ be a contractible compact domain with nonempty interior and v be a vector field on $\bar{\mathcal{T}}$. Then $\partial v_k/\partial t_l = \partial v_l/\partial t_k$ for all $k, l \in \{1, \ldots, m\}$ with $k \neq l$, if and only if there exists a twice continuously differentiable potential $u : \mathcal{T} \to \mathbb{R}$, i.e.,*

$$v(t) = \nabla_t u(t),$$

for all $t \in \mathcal{T}$.

Proof. See Zorich (2004, Vol. II, p. 296).

2.5 Applications

2.5.1 Optimal Capacity Choice and Product Distribution

Suppose that a firm can choose the quantity (or, equivalently, the quality) q of a product that will be provided to each of m geographically dispersed consumers. Each consumer $k \in \{1, \ldots, m\}$ is located at a point $t_k \in \mathbb{R}$, representing his or her type. To deliver the product to consumer i the firm incurs a quadratic transportation (or, customization) cost $d(z - t_k)^2$, where $z \in \mathbb{R}$ is the location of a distribution center that the firm is able to freely determine and d is a positive constant. The firm's unit transportation cost from its factory (located at the origin) to the distribution center is given by the smooth convex function $C(z)$ with $C'(0) = C(0) = 0$. To keep our analysis simple, we assume that each consumer's demand can be represented by a linear inverse demand function of the form $a - bq$, where a, b are positive constants.[25]

[25] This corresponds to a quadratic utility function $u_k(q) = aq - bq^2/2$ for all consumers $k \in \{1, \ldots, m\}$. Allowing for demand heterogeneity with $u_k(q) = a_k q - b_k q^2/2$ leads to analogous results.

The firm's profit can therefore be written as

$$\Pi(q,z,t) = mq(a-bq) - dq\sum_{k=1}^{m}(t_k - z)^2 - mqC(z), \tag{20}$$

where $t = (t_1, \dots t_m) \in \mathbb{R}^m$. The firm thus solves the parameterized optimization problem

$$\max_{(q,z)\in\mathbb{R}_+\times\mathbb{R}} \Pi(q,z,t). \tag{21}$$

Provided that a is sufficiently large, it is easy to verify that problem (21) has a unique interior solution $(q^*(t), z^*(t))$ and satisfies Assumptions 1–3. The firm's MCS problem is: *how does the optimal per-consumer production quantity q^* vary with t?* Starting from a market characterized by the parameter vector $t \in \mathbb{R}^m$, *is there a direction in the parameter space in which the optimal production quantity q^* increases?* We first determine the first-order necessary optimality conditions for (21),

$$\frac{\partial \Pi}{\partial q} = a - 2bq - d\sum_{k=1}^{m}(t_k - z)^2 - mC(z) = 0, \tag{22}$$

and

$$\frac{\partial \Pi}{\partial z} = 2dq\sum_{k=1}^{m}(t_k - z) - mqC'(z) = 0. \tag{23}$$

Therefore, we can restrict $\mathcal{R}(t)$ to the subset of tuples $(q,z) \in \mathcal{X} = \mathbb{R}_+ \times \mathbb{R}$ that satisfy (22) and (23). To apply the method, we also compute Φ, H^{-1} and K. Since $\varphi(q,z) = q$, we have that $\Phi = (1,0)$ and Assumption 4 is satisfied. Moreover,

$$H = \begin{bmatrix} -2b & 2d\sum_{k=1}^{m}(t_k - z) - mC'(z) \\ 2d\sum_{k=1}^{m}(t_k - z) - mC'(z) & -2mdq - mqC''(z) \end{bmatrix}.$$

The first-order necessary optimality condition with respect to z, given in (23), simplifies the Hessian matrix H to

$$H = \begin{bmatrix} -2b & 0 \\ 0 & -2mdq - mqC''(z) \end{bmatrix}.$$

It follows that[26]

$$H^{-1} = \frac{1}{2bmq(2d + C''(z))} \begin{bmatrix} -2mdq - mqC''(z) & 0 \\ 0 & -2b \end{bmatrix},$$

whence

$$-\Phi H^{-1} = \left(\frac{1}{2b}, 0\right).$$

The first row of K is given by[27]

$$\left(\frac{\partial^2 \Pi}{\partial q \partial t_1}, \ldots, \frac{\partial^2 \Pi}{\partial q \partial t_m}\right) = 2d\left((z - t_1), \ldots, (z - t_m)\right).$$

Therefore, in order to satisfy Assumption 5, we are looking for a vector $v(t) \in \mathbb{R}^m$ such that

$$\frac{d}{b}\langle (z - t_1, \ldots, z - t_m), v(t) \rangle \geq 0,$$

for all $z \in \mathbb{R}$, or equivalently

$$\langle (ze - t), v(t) \rangle \geq 0, \qquad (24)$$

[26]Observe that the determinant is nonzero, since $C''(z) \geq 0$ by convexity of g.

[27]There is no need to compute the second row, since K is left-multiplied by $-\Phi H^{-1}$, whose second component is zero.

for all $z \in \mathbb{R}$, where $e = (1, \ldots, 1)/m$ is the unit vector of the first bisectrix[28] Δ in $\mathbb{R}^m$. It is easy to see that if $v(t)$ is orthogonal to e and has a nonnegative scalar product with $-t$, the condition is satisfied. The vector $v(t) = -t + m\langle t, e\rangle e$ is such that *first*, $\langle e, v(t)\rangle = 0$, since $\langle e, e\rangle = 1/m$; and *second*, $\langle -t, v(t)\rangle = \langle t, t\rangle - m(\langle t, e\rangle)^2 \geq 0$ by the Cauchy-Schwarz inequality. Moreover, the inequality is strict if t is not collinear with e. Last, observe that, when seen from t, $v(t)$ points directly[29] to the first bisectrix Δ of $\mathbb{R}^m$. We therefore conclude from Theorem 1 that $q^*(t)$ increases as t gets closer to Δ. In other words, as the consumer types become "closer," the optimal product quantity increases. When the consumer types are identical (t collinear to e), the optimal production reaches its maximum. The problem can thus be reparameterized in the following way: define a cylinder $\mathcal{P}$ around the first bisectrix Δ, for example

$$\mathcal{P} = \left\{t \in \mathbb{R}^{m-1} : d(t, \Delta) = 1\right\},$$

where d is the Euclidian distance from a point to a line. This cylinder is an $(m-1)$-dimensional manifold, which can be parameterized by $m-1$ components. Moreover, $\mathcal{P}$ is transverse to all trajectories, and is hit by all trajectories once, so that Assumption 6 is satisfied.[30] Therefore, we have a global reparameterization of $\mathbb{R}^m$ where $m-1$ components correspond to the position on the cylinder and determine a radius emanating from Δ, and the remaining component is a parametric representation of the radius. In this particular context, it is possible to construct a more efficient parameterization: let $\mathcal{H}$ denote the $(m-1)$-dimensional hyperplane of $\mathbb{R}^m$ orthogonal to the first bisectrix and going through the origin, and $(\epsilon_1, \epsilon_2, \ldots, \epsilon_{m-1})$ be an orthogonal

[28] The first bisectrix is defined by the equation $t_1 = t_2 = \ldots = t_m$.

[29] The vector $v(t)$ points in the direction of the orthogonal projection of t on Δ.

[30] To be rigorous, $\mathcal{P}$ is diffeomorphic to $\mathcal{S}^{m-2} \times \mathbb{R}$, where $\mathcal{S}^{m-2}$ is the unit sphere in $\mathbb{R}^{m-1}$. This parameterization is a generalization of cylindric coordinates in $\mathbb{R}^3$.

basis of $\mathcal{H}$. Then, $(e_1, e_2, \ldots, e_m) = (\epsilon_1, \epsilon_2, \ldots, \epsilon_{m-1}, e)$ is an orthogonal basis of $\mathbb{R}^m$. Moreover, if t is represented with respect to that basis, i.e., $t = s_1e_1+s_2e_2+\cdots+s_me_m$, we obtain a new parameterization of the parameter space such that $q^*(s_1, \ldots, s_m)$ is nonincreasing in $(s_1, \ldots, s_{m-1})$ (the smaller these coordinates, the closer is t to the first bisectrix).[31]

It is worth observing that in order to solve the MCS problem, we relied on our knowledge of the first-order optimality condition for z. It is our second use of the crucial device $\mathcal{R}(t)$, which enables us to narrow down the domain on which Assumption 5 must be satisfied. On the other hand, we did not use the fact that the optimizer (q^*, z^*) satisfies the first-order optimality condition for q. Thus $\mathcal{R}(t)$ could have been larger without affecting our ability to construct the vector field v satisfying Assumption 5. We also note that classic supermodularity is of no use in this problem, since q^* is not monotonic in any of the t_k's. Last, observe a remarkable fact in our analysis of this example: we are able to obtain monotone comparative statics for q^* without solving explicitly for either q^* or z^*. In general, the method can be used to derive monotone comparative statics for any single decision variable, say, x_1, while one is able to solve the optimization problem explicitly for some other variables, say, $x_k(t), x_{i+1}(t), \ldots, x_n(t)$. In that case the reduced feasible set $\mathcal{R}(t)$ can be narrowed down to the set of all $x \in \mathcal{X}$ such that $x_k = x_k(t), \ldots, x_n = x_n(t)$.

[31]Moreover, it can be shown that $q^*(s)$ is independent of the last component, s_m.

2.5.2 Neoclassical Production

Consider a firm's optimal choice of factor inputs, capital k and labor l, so as to maximize the objective function

$$f(x,t) = g(k,l) - rk - wl, \tag{25}$$

where $x = (k,l)$ and $t = (r,w)$ with r the rate of return of capital and w the average wage rate. As pointed out by Milgrom and Shannon (1994), if g is not supermodular, comparative statics are not monotone in the original parameterization. To demonstrate the use of our method we assume that g is twice continuously differentiable and that there exists a unique optimizer in the interior of $\mathbb{R}^2_+$. Therefore, Assumptions 1–3 are satisfied, with the Hessian and cross-derivative matrices

$$H = \begin{bmatrix} g_{kk} & g_{kl} \\ g_{kl} & g_{ll} \end{bmatrix} \text{ and } K = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

respectively. The pseudo-gradient is therefore

$$W = \frac{1}{D} \begin{bmatrix} g_{ll} & -g_{kl} \\ -g_{kl} & g_{kk} \end{bmatrix},$$

where $D(k,l) = (g_{kk}g_{ll} - g_{kl}^2)(k,l)$ is the determinant of $H(k,l)$. Because of the strict concavity of g at the optimizer, we can restrict the reduced feasible set $\mathcal{R}(t)$ to the subset of $\mathbb{R}^2_+$ where H is negative definite, implying that $D(k,l)$ is positive.[32] Notice that the pseudo-gradient and the reduced feasible set are independent of r and w.[33]

[32]The determinant D is positive at any maximizer of (25) as the product of the two negative eigenvalues of H.

[33]Note that H,K and D are all independent of (r,w).

To simplify our exposition we drop the explicit dependence on t and refer to $\mathcal{R}(t)$ as $\mathcal{R}$. If one can find a vector v making a positive scalar product with $W(k,l)$ for all (k,l), it will satisfy Assumption 5 for all values of r and w. The vector field will then consist of straight, parallel trajectories of direction v. This will generate a linear reparameterization of the problem (the basis of the new coordinate system consisting of v and any other vector not collinear to v), under which both k and l are nondecreasing in the first parameter coordinate. Before addressing the problem of monotone comparative statics for k and l at the same time, let us consider the simpler problem of finding monotone comparative statics for k alone. That is, we consider the function $\varphi(k,l) = k$, which trivially satisfies Assumption 4. The pseudo-gradient then becomes

$$W_k(k,l) = \Phi(k,l)W(k,l),$$

with $\Phi(k,l) = (1,0)$. This yields

$$W_k(k,l) = \frac{1}{D}(g_{ll}, -g_{kl}).$$

We are looking for a vector $v \in \mathbb{R}^2 \setminus \{0\}$ such that $\langle W_k(k,l), v\rangle \geq 0$, or equivalently

$$\langle (g_{ll}, -g_{kl}), v\rangle \geq 0. \tag{26}$$

Since g_{ll} is nonpositive, a solution is $v = (-1,0)$. That is, $k(r,w)$ is nondecreasing in r. Notice that this result obtains without assumption on g except for smoothness.[34] In general, $k(r,w)$ is not monotonic in w: this would require $g_{kl} \geq 0$ for all (k,l) (as

[34] Another way to see this is the following: the function $f(k,l,r,w) = g(k,l) - rk - wl$ is supermodular in (r,k). To apply standard supermodularity results, define $F(k,r,w) = \max_{l \geq 0} f(k,l,r,w)$. F is supermodular in (k,r) and $\arg\max_{k \geq 0} F(k,r,w) = k(r,w)$. This implies that $k(r,w)$ is nondecreasing in r. We thank Paul Milgrom for this observation.

can be seen by substituting $v = (0, -1)$ in equation (26)). However, there may be other directions of v such that k is nondecreasing. Equation (26) can be rewritten as

$$v_1 g_{ll}(k,l) - v_2 g_{kl}(k,l) \geq 0$$

for all $(k,l) \in \mathbb{R}^2_+$. Since $g_{ll} \leq 0$ on $\mathcal{R}$, this is equivalent to

$$v_1 \leq \min\{D, d\}\, v_2 \tag{27}$$

where $(d, D) = (\inf \delta, \sup \Delta)$, with

$$\begin{aligned} \delta &= \left\{ \frac{g_{kl}}{g_{ll}}(k,l) : g_{kl} \geq 0,\ g_{ll} < 0,\ (k,l) \in \mathcal{R} \right\}, \\ \Delta &= \left\{ \frac{g_{kl}}{g_{ll}}(k,l) : g_{kl} \leq 0,\ g_{ll} < 0,\ (k,l) \in \mathcal{R} \right\}, \end{aligned}$$

as well as the conventions that $\inf\{\emptyset\} = +\infty$ and $\sup\{\emptyset\} = -\infty$. When $\Delta \neq \emptyset$, $D \geq 0$. Similarly, $d \leq 0$ if $\delta \neq \emptyset$. When g is supermodular, Δ is empty or reduced to the singleton $\{0\}$, so that $D \leq 0$. Moreover, $\delta \neq \emptyset$ implies $d \leq 0$, so that condition (27) is satisfied by any $v \in \mathbb{R}^2_-$, by virtue of the nonpositivity of $d \wedge D$. This proves that $k(r, w)$ is nonincreasing not only in r, but also in w, whenever g is supermodular. In general, relation (27) defines a convex cone $\Gamma_k \in \mathbb{R}^2$ based at the origin which always contains the negative real line $\mathbb{R}_- \times \{0\}$. Except when both $D = +\infty$ and $d = -\infty$, Γ_k has a nonempty interior. If $g_{kl} > 0$, $D = -\infty$, implying that Γ_k is a half-space that is located below the line $v_1 = d\, v_2$.

The optimizer $k(r, w)$ is nondecreasing in any direction of Γ_k. When Γ_k has a nonempty interior, it is possible to change coordinates in the parameter space by using two independent basis vectors in Γ_k. As pointed out earlier, this coordinate

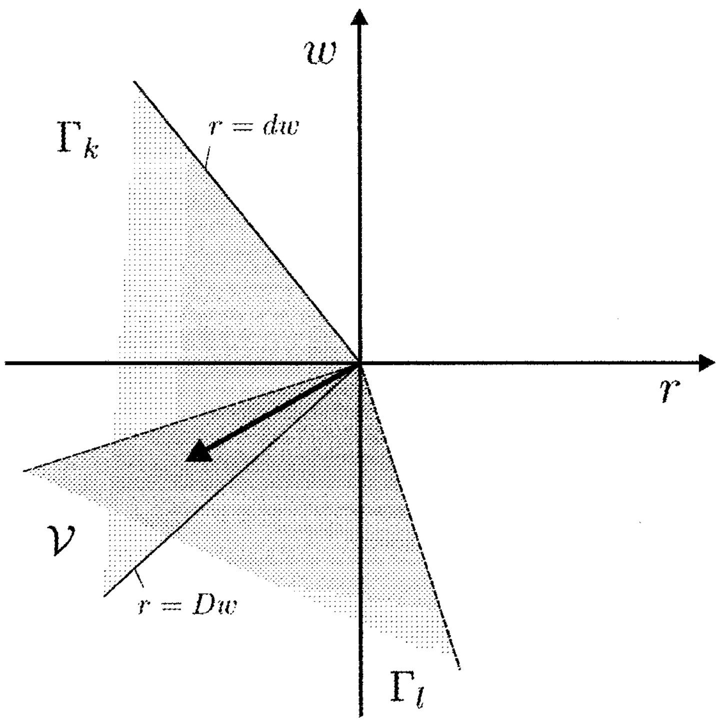

Figure 4: Neoclassical Production: Monotonicity of $k(r, w)$ on $\Gamma_k \cap \Gamma_l$.

change is global, since Γ_k is independent of the particular values chosen for r and w. Similarly, $l(r, w)$ is nondecreasing in any direction located in the cone Γ_l based at the origin and containing the negative imaginary line $\{0\} \times \mathbb{R}_-$. Having constructed these two cones we can now address the more challenging question, *is it possible to find directions in which both k and l increase?* The answer depends on $\mathcal{V} = \Gamma_k \cap \Gamma_l$. If $\mathcal{V}$ is empty, we cannot construct any direction that jointly increases k and l. If $\mathcal{V}$ is nonempty, then it is also a convex cone, whose elements are directions of joint increase. The intersection $\mathcal{V}$ being empty does not prove the nonexistence of directions of joint increase. It just means that we do not have enough information on the optimizers to produce such directions. As our information gets richer, the set $\mathcal{R}$ becomes narrower, which implies that the cones Γ_k and Γ_l become wider. When one has enough information, the cones are wide enough to intersect, yielding the desired directions of joint increase (cf. Figure 4).[35] When $\mathcal{V}$ is nonempty, the vector field can be chosen constant: $v(r, w) = v$ for some $v \in \mathcal{V}$. In that case, any straight line $\mathcal{P}$ orthogonal to v satisfies Assumption 6: it is transverse and is hit exactly once by all trajectories. The reparameterization is then simple: take any vector e_2 on that line and let $e_1 = v$. Then, (e_1, e_2) is an orthogonal basis of $\mathbb{R}^2$, such that if one expresses $t = (r, w)$ on that basis (that is, $(r, w) = s_1 e_1 + s_2 e_2$), then $(k(s), l(s))$ is nondecreasing in s_1.

[35]The analysis for obtaining directions of joint decrease is naturally analogous, by considering opposite directions. Similarly, it is possible to obtain directions of increase in one parameter and decrease in the other.

2.5.3 Giffen Goods

In an economy with two goods, an agent wishes to maximize her increasing and concave utility by solving

$$\max_{(x,y)\in\mathbb{R}^2_+} u(x,y),$$

subject to

$$px + qy \leq w,$$

where x, y are the quantities of the two goods, p and q are their respective positive prices, and w represents the agent's wealth. Without loss of generality, we select the second good to be the numéraire and correspondingly set $q = 1$. In addition, since both goods are desirable, the agent's budget constraint is binding. As pointed out in Section 2.3.5, the agent's problem can then be restated as

$$\max_{x\in[0,w/p]} u(x, w - px).$$

In this formulation the problem has one decision variable and two parameters.[36] If u is smooth, compactness insures the existence of an optimizer. We assume that for our starting values of w and p, the optimizer $x(w,p)$ is unique, located in $(0, \frac{w}{p})$, and that $u(\cdot, w, p)$ is locally strictly concave at $x(w,p)$ and everywhere twice continuously differentiable. This implies that Assumptions 1–3 are satisfied. The first good is "normal" if $x(w,p)$ is nonincreasing in p, and Giffen or "inferior" if this monotonicity is sometimes violated.[37] Similarly, one would usually expect the consumption of

[36] As mentioned earlier, we have therefore converted the initial problem with two decision variables on a one-dimensional manifold of $\mathbb{R}^2$ with empty interior but nonempty relative interior, into a problem with one decision variable on a set with nonempty interior.

[37] Classic examples include potatoes or bread. The gist of the argument goes as follows: when the price for bread increases, poorer people cannot afford buying "luxury goods" such as meat, and end up consuming more bread, which is still the cheapest good. Other goods violating this monotonicity

any good to increase with the agent's wealth. However, this monotonicity is also sometimes violated. The question then becomes, *under what conditions is a good normal, and how are price and wealth effects connected?* The Hessian and cross-derivative matrices are

$$H = u_{xx} - 2pu_{xy} + p^2 u_{yy} \quad \text{and} \quad K = [u_{xy} - pu_{yy} \quad - xu_{xy} + pxu_{yy} - u_y] \, .$$

The pseudo-gradient W is therefore given by

$$W(x,w,p) = \frac{1}{D} [\alpha(x,w,p) \quad - x\alpha - u_y(x, w - xp)] \, ,$$

where $D = -H > 0$ (by strict concavity of H at the optimizer) and $\alpha(x,w,p) = u_{xy} - pu_yy$. We also note that since $\varphi(x) = x$, Assumption 4 is trivially satisfied. In order to meet Assumption 5, we are thus looking for a vector $v \in \mathbb{R}^2$ such that

$$\alpha(x,w,p)v_1 - (x\alpha(x,w,p) + u_y)v_2 \geq 0 \tag{28}$$

for all x in the reduced feasible set $\mathcal{R}(w,p)$. First, we observe that if u is supermodular and concave in its second variable, α is nonnegative, which implies, along with the nonnegativity of u_y, that any vector v in $\mathbb{R}_+ \times \mathbb{R}_-$ solves (28). This means that if the two goods are complements and if the utility function is concave in the second good, then the first good is normal.[38] In the general case, we show that there is a hierarchical relationship between wealth and price effects. Increasing wealth amounts to setting $v_1 > 0$ and $v_2 = 0$, so that the good is normal with respect to the wealth

are Veblen goods (Veblen, 1899) or positional goods (Hirsch, 1976, Chapter 3) which are such that the implied status of the owner increases with their price.

[38]This result can also be shown by observing that the concavity of u in y implies the supermodularity of v in $(x,w,-p)$.

effect if and only if $\alpha \geq 0$. On the other hand, $\alpha \geq 0$ implies that $x\alpha + u_y \geq 0$. Since decreasing the price amounts to setting $v_2 > 0$ and $v_1 = 0$, the good is therefore normal with respect to the price effect if $\alpha \geq 0$. This shows that the following result holds for any smooth, nondecreasing utility function: if an augmentation in wealth increases the optimal consumption of a good, then a cut in its price also increases its optimal consumption. The reverse is, in general, not true.[39] If the optimal consumption $x(w,p)$ is known or constrained to belong to some subinterval $\mathcal{J} = (x_1, x_2) \subset (0, w/p)$,[40] the analysis can be refined. For example, suppose that $\min_{x \in \mathcal{J}} \{\alpha(x,w,p)\} \geq 0$ for all w, p. Then the good is normal with respect to both wealth and price effects (any $v \in \mathbb{R}_+ \times \mathbb{R}_-$ solves (28)). If α sometimes takes negative values but $\min_{x \in \mathcal{J}} \{x\alpha(x,w,p) + u_y(x, w - px)\} \geq 0$, then the good is normal with respect to price effects ($v \in \{0\} \times \mathbb{R}_-$ solves (28)). More generally, suppose that $\alpha(\cdot, w, p)$ changes sign only once on (x_1, x_2), and that $\alpha(x_1, w, p) > 0$ for all w and p in an open neighborhood of initial values of wealth and price. The second condition means that the good is normal for low consumption, while the first condition means that the good becomes Giffen for high consumption values. Then, if the vectors

$$(\alpha(x_1, w, p), -x_1\alpha(x_1, w, p) + u_y(x_1, w - px_1))$$

and

$$(\alpha(x_2, w, p), -x_2\alpha(x_2, w, p) + u_y(x_2, w - px_2))$$

are in the same half-plane, there exists a normal vector $v(w,p)$ of the half-space whose scalar product with $W(x,w,p)$ is nonnegative for all $x \in (x_1, x_2)$. The situation is

[39]The result can be read in the opposite direction: if a good is inferior with respect to price effect, it is also inferior for wealth effect.

[40]For example, minimal consumption could be imposed or supply could be limited.

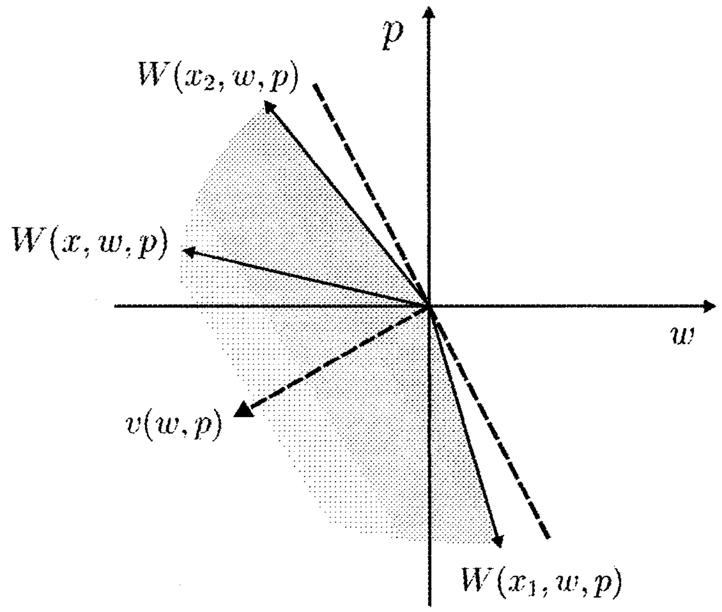

Figure 5: Giffen Goods: Solving the MCS Problem for $x(w, p)$.

represented in Figure 5. It is easy to verify that $v(w, p)$ can always be taken in the negative orthant $\mathbb{R}^2_-$. Therefore we have the following result: if the pseudo-gradients of consumption boundaries x_1, x_2 lie in the same half-plane, and if the good behaves as a normal good for low consumption values and as a Giffen one for high consumption values (such as potatoes, cf. footnote 37), then there exists a way to increase optimal consumption this good by reducing both wealth and price at the same time. It can also be shown that in the same situation, there is no way to increase consumption by raising both wealth and price at the same time.

2.5.4 Multiattribute Screening

A variation of our method can be used in the context of screening with multiple instruments. A firm faces customers of different types, distributed on an interval $\mathcal{X} \subset \mathbb{R}$ according to a positive density function g. The firm sells products whose attributes are described by a vector $t \in \mathcal{T} \subset \mathbb{R}^m$. The goal of the firm is to propose a product line $\Gamma \subset \mathcal{T}$ and a price schedule $P : \Gamma \to \mathbb{R}_+$ (with $0 \in \Gamma$ and $P(0) = 0$) that maximizes its expected profit

$$\pi(P, \Gamma) = \int_{\mathcal{X}} [P(t(x)) - C(t(x))]\, g(x)dx,$$

where $C(t(x))$ is the cost of producing $t(x) \in \Gamma$ and $t(x)$ solves the type-x consumer's utility maximization problem

$$t(x) \in \arg\max_{t \in \Gamma} \{u(x,t) - P(t)\}.$$

This general screening problem has only been solved in particular cases[41] When $m = 1$ (only one instrument available), it is possible to directly compute the optimal price schedule $P(x)$ under some additional supermodularity assumptions on the primitives of the problem. Assuming that u is smooth, we can define

$$\mu(x,t) = u(x,t) - u_x(x,t)\frac{1 - G(x)}{g(x)},$$

where G is the probability distribution of the density g. Mussa and Rosen (1978) have shown, based on a technique developed by Mirrlees (1971), that if u and μ are twice differentiable and supermodular (i.e., $u_{xt} \geq 0$ and $\mu_{xt} \geq 0$ on $\mathcal{X} \times \mathcal{T}$), then it is possible to construct the optimal price schedule. Moreover, this optimal schedule leads to "perfect" screening (without bunching): each consumer of type x will buy a distinct product $t(x)$. Suppose now that there are $m \geq 2$ product attributes available.

[41]See Roberts (1979), Mirman and Sibley (1980), and Matthews and Moore (1987) for the multi-attribute, one-dimensional type case, and Rochet and Stole (2003) for a recent account of the general multidimensional screening literature.

Our method can be extended to build product lines that perfectly screen consumers. Defining the pseudo-gradient W as the $(2 \times m)-$matrix

$$W(x,t) = \begin{bmatrix} \nabla_x^T u_t(x,t) \\ \nabla_x^T \mu_t(x,t) \end{bmatrix},$$

suppose that there exists for all $t \in \mathcal{T}$ a nonzero vector $v(t)$ such that

$$W(x,t)v(t) \geq 0$$

for all $x \in \mathcal{X}$ (i.e., Assumption 5 is satisfied). We can then define $\Gamma \subset \mathcal{T}$ to be the image of any smooth trajectory $\gamma : (0,1) \to \mathcal{T}$ generated by the vector field v. This leads to a reparameterization of the utility u and the function μ when restricted to $\mathcal{X} \times \Gamma$ (that is, when customers are offered the product line Γ). Specifically, we define $\tilde{u}$ and $\tilde{\mu}$ on $\mathcal{X} \times (0,1)$ by $\tilde{u}(x,\lambda) = u(x,\gamma(\lambda))$ and $\tilde{\mu}(x,\lambda) = \mu(x,\gamma(\lambda))$. Using Lemma 1 and Theorem 1, we can show that $\tilde{u}$ and $\tilde{\mu}$ are supermodular on $\mathcal{X} \times (0,1)$. The aforementioned result then implies that it is possible to find the optimal price schedule on Γ, and that this schedule perfectly screens customers. This approach does not solve the original problem of maximizing the profit on $\mathcal{T}$, since we artificially restricted ourselves to the product line Γ. However, the method can be repeated for several distinct trajectories, and leads to a perfectly screening price schedule that maximizes the expected profit not only on a particular product line, but on a large class of product lines that spans the whole multiattribute space $\mathcal{T}$.

2.6 Discussion

In the available literature on monotone comparative statics, the parameterization of the optimization problem is essentially taken as given.[42] The presently known criteria for the monotonicity of solutions hold, therefore, only with respect to the particular problem formulation given at the outset. Milgrom and Shannon's (1994) characterization of the monotonicity of solutions to (1) on lattices requires the objective function f to be quasi-supermodular in x and to satisfy a single-crossing property in (x, t).[43] The supermodularity requirement on the objective function can thereby be interpreted in terms of "complementarity" of decision variables, a concept that dates back at least to Edgeworth (1897) and whose origins are reviewed by Samuelson (1974). Milgrom and Roberts (1990) demonstrate the power of complementarities and associated supermodularity properties in interpreting decision changes as monotone responses to exogenous shifts of economic conditions. Even though equilibria cannot be located exactly, complementarities allow one to make precise statements about the direction in which optimal decisions change as a consequence of parameter changes. In the absence of such complementarities, the presently available theory unfortunately guarantees the non-monotonicity of solutions, even though this non-monotonicity might just be a symptom of an unsuitable parameterization of the problem. The chapter proposes a way to obtain an equivalent formulation of the optimization problem (1) using a new parameterization, such that – provided sufficient knowledge about the location of the solution – monotone comparative statics may be obtained. Finding a new parameterization of the problem amounts to creating a set of

[42]Note that the decision variables are also typically taken as given. Our method in principle allows for a change of the decision variables to obtain monotone comparative statics through an appropriate choice of the evaluation function φ.

[43]Athey (2002) applies these results to expected-utility maximization problems under uncertainty and finds necessary and sufficient conditions on the model primitives in that context.

economic indicators which allow for monotonic decision making and thus easy rules of thumb (i.e., when the relevant indicator goes up, the optimal decision goes up, too). This seems especially useful in situations where the same optimization problem needs to be solved repeatedly for different parameter values. Let us briefly mention at this point that our method naturally extends to equilibrium problems (cf. also Milgrom and Roberts (1994)) specified by a relation

$$F(x, t) = 0,$$

where $F : \mathcal{X} \times \mathcal{T} \to \mathbb{R}^n$ is a continuously differentiable function, as can be seen by associating $\nabla_x f$ with F, so that $H = \nabla_x F$ and $K = \nabla_t F$. With these substitutions in place, all of our results hold essentially without any change. We also remind the reader that, as pointed out in Section 2.3.5, even though we require through Assumption 2 the existence of a unique interior global optimum, parameter-dependent constraints can be accommodated in a straightforward way by shifting the analysis to a submanifold in $\mathcal{X}$ or by augmenting the space of decision variables by Lagrange multipliers corresponding to the binding constraints.

Sometimes our method may also be useful for reducing the number of parameters without any losses. To show this, let us first note that, clearly, an "ideal" parameterization of problem (1) for the case $m \geq n$ associates exactly one parameter t_i with each component of the decision variable x_i, and is such that x_i remains unchanged in response to a change of parameter t_j (with $j \neq i$). If the location of the optimal solution is perfectly known, a reparameterization with these "ideal" properties can always be obtained by setting $s_i = x_i(t_i)$ for $i \in \{1, \ldots, n\}$ and simply discarding all other $m - n$ parameters, $t_{n+1}, \ldots, t_m$. Unfortunately, lack of knowledge about the location of the optimizer (up to a monotone transformation) usually makes this trivial solution impracticable. Nevertheless, it may sometimes be possible to reduce the number of parameters by finding directions $v(t)$ for which $\langle W(x,t), v(t)\rangle = 0$ for all x in an admissible reduced feasible set $\mathcal{R}(t)$. If such a direction can be found, the solution does not depend on λ in the associated global MCS reparameterization and λ may thus be discarded from the set of new parameters. By repeating this process it may be possible to eliminate further parameters.

The tradeoff between the decision maker's knowledge about the location of the optimal action and her ability to find directions that guarantee monotone behavior of the optimizer (i.e., to solve the (local) MCS problem) is related to "partially specified problems", as discussed by Milgrom (1994). One type of partially specified optimization problems possesses an objective function of the form $f(x,t) = g(x,t) + \delta(x)$, where δ is any affine mapping from $\mathcal{X} \subset \mathbb{R}$ to $\mathbb{R}$. Monotonicity of optimal solutions $x_\delta(t)$ to the problem (1) for *any* affine δ is then equivalent to the (otherwise unknown) function g being supermodular on $\mathcal{X} \times \mathcal{T}$. The key idea in this approach is that the class of perturbations δ is large enough relative to g and $\mathcal{X}$ to allow for any location of the optimizer in $\mathcal{X}$. A variation of our method bypasses this definitive result when the function g fails to be supermodular: it might be possible to transform the parameter space so as to "supermodularize" the function g. If $g(x,t)$ is not supermodular in (x,t), we can build trajectories $\gamma : (0,1) \to \mathcal{T}$ in the parameter space such that $g(x, \gamma(\lambda))$ is supermodular in (x, λ). We have also used this approach in Section 2.5.4 and Corollary 2. While supermodularization of functions is just a particular application of our method, its repeated use in this chapter suggests its potential benefits in numerous other settings, such as for supermodularizing noncooperative games.

Further research could proceed to relax some of the differentiable structure imposed to obtain our results. Systematic MCS reparameterizations can be expected to naturally generalize to an analysis on lattices. The problem is to find a rule on $\mathcal{W} \subset \mathcal{T} \times \mathcal{T}$ such that $(t, t') \in \mathcal{W}$ implies $\phi(x(t')) \geq \phi(x(t))$. In particular, suppose that we can build trajectories $\{\mathcal{T}_i\}$ in $\mathcal{T}$ such that $(x, t) \mapsto f(x, t)$ has the single-crossing property[44] on $\mathcal{X} \times \mathcal{T}_i$ for all $i \in \mathcal{I}$. If f is in addition (quasi-)supermodular in x, standard results as in Milgrom and Shannon (1994) apply, showing that $x(t)$ is nondecreasing along the trajectories. The problem is of course to construct such trajectories. Our method for doing so is based on differential calculus, but there may be other ways to build trajectories, or at least to find rules in the parameter space, in order to achieve monotone comparative statics (e.g., through discretization of our results).

[44]Any trajectory is totally ordered, with the order implied by the parametric description of the curve.

Chapter 3

Games Played Through Agents

For the game of complete information with multiple principals and multiple common agents discussed by Prat and Rustichini (2003), we construct a general set of equilibrium transfers that implement any efficient outcome as a weakly truthful equilibrium, and the subset of such equilibria that are Pareto optimal for the principals. We provide conditions under which the general set completely characterizes the set of all weakly truthful equilibria implementing a given efficient outcome. We also show that under the sole assumption of concave payoff functions, any efficient outcome can be implemented with principals using affine strategies.

3.1 Introduction

Recently, Prat and Rustichini (2003) [P&R] analyzed a class of games of complete information in which N common agents noncooperatively implement an outcome (i.e., a collection of actions) after having received offers of outcome-contingent transfers by M principals. Based on Bernheim and Whinston's (1986) truthful Nash equilibrium for games of common agency with one principal and N agents, P&R introduce weakly truthful equilibrium (WTE) as a Nash-equilibrium refinement for their class of *games played through agents.* They focus on the implementation of an *efficient outcome* (that maximizes the sum of all principals' and agents' payoffs). P&R provide a characterization of WTEs and prove that any WTE must be efficient. Moreover, they show that *any* efficient outcome can be implemented as a WTE when agents have convex action sets and all parties have bounded, concave, and continuous payoff functions (P&R, Theorem 8). Their existence proof, based on a generalization of Farkas' Lemma (Aubin and Ekeland, 1984, p. 144), is nonconstructive. Hardly any insight is gained about *how* weakly truthful equilibria can actually be implemented, i.e., which transfers to specify in the equilibrium contracts. As Weber and Xiong (2004) [W&X] demonstrate, it is precisely the latter question of equilibrium implementation which is of great importance in practical applications such as the coordination of supply chains. Indeed, as off-equilibrium payoffs supporting the implementation of an efficient outcome vary, in-equilibrium payoffs to principals and agents are reallocated. We provide several direct algorithms for implementing any efficient outcome as a WTE in a game played through agents, under the assumption that all principals' and agents' payoff functions are concave and continuous.[1]

Given any efficient outcome of a game played through agents, this chapter pursues the following *three main objectives*, corresponding to the underlying practical contract design problem: *first,* to construct a general set of (and, whenever possible, the set of *all*) WTEs implementing the efficient outcome and to provide a simple representation of these WTEs; *second,* to characterize the subset of these WTEs that yield Pareto-optimal in-equilibrium transfers for the principals; *third,* given any attainable

[1]While these properties require assumptions on the domains of the payoff functions, our results are valid when the actions sets are finite, or more generally are any subsets of these domains.

Pareto-optimal in-equilibrium transfer, to provide a WTE that implements the efficient outcome and results in the prescribed in-equilibrium transfers. The chapter is organized as follows. Section 3.2 describes several applications of the games played through agents. Section 3.3 reviews the setting of games played through agents and recalls a simple equilibrium characterization (Proposition 1) by P&R and W&X. This characterization reduces the problem of finding a WTE that implements a given efficient outcome to a problem of constructing separating "excess transfers." In Section 3.4 we provide an inductive algorithm (Theorem 1) to obtain separating transfers which are affine for each principal except one. In the generic case where the payoff functions are differentiable at the efficient outcome, we show how to implement the outcome with excess transfers that are affine for all principals (Proposition 2). We use the first algorithm to systematically find additional "maximal" excess transfers through outcome-contingent convex combination (relation (27)). Using a leveling algorithm, it is possible to further extend the set of equilibrium excess transfers such that "minimal" excess transfers appear as limits of the algorithm (Proposition 3). We are thus able to describe a general set of (and, whenever these minimal excess transfers are unique, the set of *all*) equilibrium excess transfers (Theorem 2) in terms of an extremal basis which can be directly computed from the minimal excess transfers. In Section 3.5, we describe the subset of these transfers which are Pareto optimal for the principals and the set of attainable best in-equilibrium transfers for each principal (Theorem 3). We also show how to implement any of these in-equilibrium transfers as a WTE. Section 3.6 provides an exact expression for the minimal excess transfers for a class of payoff functions. The minimal excess transfers are key in the implementation of efficient equilibria. Section 3.7 discusses and summarizes our results.

3.2 Applications of Games Played Through Agents

To illustrate the importance of obtaining explicit transfers implementing an efficient outcome, and the explicit subset of these transfers which are Pareto optimal for the agents, this section provides several applications of games played through agents. (See also Bernheim and Whinston (1986) and P&R.)

Supply Chains. By definition, supply chains are "coordinated" if the outcome maximizes the sum of the payoffs of all firms involved. Much of the literature on coordinated supply chains is focused on a two-echelon single-agent single-principal context, which already involves nonlinear contracts ("quantity-dependent pricing"), and contracts with discount across products and orders ("generalized tying"). In a multi-supplier context, there can also be provisions relative to actions for other suppliers ("exclusive dealing"). When demand is random, as studied in Bernstein and Federgruen (2005), contracts bear on expected values. Some contracts involve multiple components of actions, for example when suppliers must choose both capacities and quantities. In this case, buyers may be induced to enter "royalty schemes," such as pay-back and revenue-sharing contracts.[2] Principals can either be at the lower or at the higher echelon of the supply chain, depending on the allocation of the bargaining power. The model addresses both situations, by a simple transformation of the framework, as described in W&X (Section 2.5).

Labor Economics. When several firms bestow the right to make certain decisions to common intermediaries, they behave as principals influencing agents. For example,

[2] See Cachon (2003) and Weber and Xiong (2004).

insurance companies use common insurance brokers to sell their policies.[3]

Lobbying. In order to influence public decisions, interest groups engage in transactions (e.g. by financing electoral campaigns) with decision makers (see Dixit, Grossman, and Helpman (1997)). As pointed out by P&R, most decisions are now made collegially, which makes it necessary to consider a multi-agent setting.

Multi-Object Auctions. When bidders submit their offers to an auctioneer, they behave as principals influencing an agent, whose decision is to allocate the auctioned good (see Bernheim and Whinston (1986)). In the context of multi-object auctions with complementary goods, bidders are sometimes allowed to condition the payment of their bids to a given auctioneer on the result of another auction, in which case a multi-agent setting is required for the analysis of the game. A particular instance of such auctions (with one principal) are take-over bids, where the potential buyer conditions his purchase of the stocks (from stockholders, who are the auctioneers and agents of this game) on his obtaining a minimal percentage of the outstanding float of the stock, without which he does not gain the desired control over the targeted firm.

[3]These contracts have recently been the focused of an investigation by NY Attorney General Elliot Spitzer, on the count of possible bid rigging, and several of the nation's largest insurers admitted to the attorney general's office having payed kickbacks to get business from insurance brokers.

3.3 Games Played Through Agents

Let $\mathcal{M} = \{1, \ldots, M\}$ be the set of all principals and $\mathcal{N} = \{1, \ldots, N\}$ be the set of all agents, where $M, N \geq 1$. Each agent $n \in \mathcal{N}$ can implement an action $x_n \in \mathcal{X}_n$, where $\mathcal{X}_n$ is a nonempty compact subset of $\mathbb{R}_+^{ML}$ (for some $L \geq 1$) and $x_n = (x_n^1, \ldots, x_n^M)$. The component $x_n^m \in \mathbb{R}_+^L$ of agent n's action can be thought as a transfer of L goods and services ("actions") between agent n ("he") and principal m ("she"). Any game played through agents consists of two periods. In the first period, each principal $m \in \mathcal{M}$ proposes an outcome-contingent transfer $t_n^m \in C(\mathcal{X}_n, \mathbb{R}_+)$ to agent n.[4] In the second period, the transfer schedule $t = [t_n^m]$ is announced publicly and each agent n implements his most preferred action to obtain a respective net payoff of

$$U_n(x_n; t) = \Gamma_n(x_n) + \sum_{m \in \mathcal{M}} t_n^m(x_n), \tag{1}$$

where $\Gamma_n \in C(\mathcal{X}_n, \mathbb{R})$ is agent n's gross payoff function. Provided the outcome $x = (x_1, \ldots, x_N) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ is implemented by the agents, principal m's net payoff is

$$V^m(x; t^m) = \Pi^m(x) - \sum_{n \in \mathcal{N}} t_n^m(x_n), \tag{2}$$

where $\Pi^m \in C(\mathcal{X}, \mathbb{R})$ is principal m's gross payoff and $t^m = (t_1^m, \ldots, t_N^m)$ is her transfer. Principal m's *transfer* t^m *is weakly truthful* relative to an outcome $\hat{x} \in \mathcal{X}$ if

$$V^m(\hat{x}; t^m) = \max_{x \in \mathcal{X}} V^m(x; t^m). \tag{3}$$

[4]Given two topological spaces $\mathcal{R}$ and $\mathcal{S}$, we denote by $C(\mathcal{R}, \mathcal{S})$ the set of all continuous functions from $\mathcal{R}$ to $\mathcal{S}$.

A subgame-perfect *pure-strategy (Nash) equilibrium* of the two-period game

$$\mathcal{G} = \{\{\mathcal{M}, \mathcal{N}\}, \{V^m(\cdot), U_n(\cdot)\}, \{C(\mathcal{X}, \mathbb{R}_+^N), \mathcal{X}_n\}\}$$

is a pair $(\hat{t}, \hat{x}) \in C(\mathcal{X}, \mathbb{R}^{MN}) \times \mathcal{X}$ such that (i) for every $n \in \mathcal{N}$ given any $t \in C(\mathcal{X}, \mathbb{R}^{MN})$,

$$\hat{x}_n(t) \in \arg\max_{x_n \in \mathcal{X}_n} U_n(x_n, \hat{x}_{-n}; t), \tag{4}$$

and (ii) for every $m \in \mathcal{M}$, given $\hat{t}^{-m} \in C(\mathcal{X}, \mathbb{R}_+^{N-1})$,

$$\hat{t}^m \in \arg \max_{t^m \in C(\mathcal{X}, \mathbb{R}_+^N)} V^m(\hat{x}(t^m, \hat{t}^{-m}); t^m). \tag{5}$$

A *pure-strategy equilibrium* $(\hat{t}, \hat{x})$ *is weakly truthful* if each principal m's transfer is weakly truthful with respect to the equilibrium outcome $\hat{x}$. As P&R (Proposition 3) note, *any WTE is efficient* in the sense that the associated outcome $\hat{x}$ maximizes the joint surplus

$$W(x) = \sum_{m \in \mathcal{M}} V^m(x; t^m) + \sum_{n \in \mathcal{N}} U_n(x; t) = \sum_{m \in \mathcal{M}} \Pi^m(x) + \sum_{n \in \mathcal{N}} \Gamma_n(x_n).$$

For any given efficient outcome $\hat{x} \in \arg\max_{x \in \mathcal{X}} W(x)$, let $F^m(x) = \Pi^m(x) - \Pi^m(\hat{x})$ denote principal m's *excess payoff*, and let $G_n(x_n) = \Gamma_n(\hat{x}_n) - \Gamma_n(x_n)$ be agent n's *excess cost* relative to their respective payoffs at $\hat{x}$. Based on P&R's results, W&X characterize WTEs of the game $\mathcal{G}$ in the following compact form, which they term the *reduced contract design problem.*

Proposition 1 (Reduced Contract Design Problem) . *The pair $(\hat{t}, \hat{x})$ is a WTE of the game $\mathcal{G}$ if and only if, for all $(m, n) \in \mathcal{M} \times \mathcal{N}$,*

$$F^m - \sum_{n \in \mathcal{N}} \Delta_n^m \leq 0 \leq G_n - \sum_{m \in \mathcal{M}} \Delta_n^m, \tag{R}$$

where $\Delta_n^m(x_n) = \hat{t}_n^m(x_n) - \hat{t}_n^m(\hat{x}_n)$ is principal m's excess transfer to agent n contingent on the feasible outcome $x_n \in \mathcal{X}_n$.

In fact, any excess transfer $\Delta = [\Delta_n^m]$ that solves (R) for a given efficient outcome $\hat{x} \in \mathcal{X}$ (which we refer to as reduced (equilibrium) transfer) can be mapped to an admissible equilibrium transfer $t \in C(\mathcal{X}, \mathbb{R}_+^{MN})$ by setting $\hat{t}_n^m(x_n) = \hat{\Delta}_n^m(x_n) + \alpha_n^m$. The nonnegative constants α_n^m ("in-equilibrium transfers") correspond to the equilibrium transfers contingent on the implemented efficient equilibrium outcome $\hat{x}$, i.e., $[\alpha_n^m] = \hat{t}(\hat{x})$. For instance, by using for any fixed $n \in \mathcal{N}$ the recursive construction suggested by P&R (Lemma 2), one obtains

$$\alpha_n^m = -\min_{x_n \in \mathcal{X}_n} \left\{ G_n(x_n) - \sum_{i=1}^{m-1} \hat{\Delta}_n^i(x_n) - \sum_{i=m+1}^{M} \Delta_n^i(x_n) \right\} \geq 0 \tag{6}$$

and

$$\hat{\Delta}_n^m(x_n) = -\min \left\{ \alpha_n^m, -\Delta_n^m(x_n) \right\}, \tag{7}$$

for all $m = 1, \ldots, M$. More generally, if $\hat{\Delta}$ is an "admissible" equilibrium excess transfer (e.g., obtained through (6)–(7)), then the in-equilibrium transfers α_n^m can be obtained through[5]

$$\alpha_n^m = -\min_{x_n \in \mathcal{X}_n} \left\{ \hat{\Delta}_n^m(x_n) + \left(G_n(x_n) - \sum_{i \in \mathcal{M}} \hat{\Delta}_n^i(x_n) \right) \right\} \geq 0 \tag{8}$$

for all $(m,n) \in \mathcal{M} \times \mathcal{N}$. The continuous equilibrium excess transfer $\hat{\Delta} = [\hat{\Delta}_n^m]$ is thereby *admissible* if it solves the reduced contract design problem (R) *and* if it is such that

$$\hat{t}_n^m = \hat{\Delta}_n^m + \alpha_n^m \geq 0 \tag{9}$$

for all $(m,n) \in \mathcal{M} \times \mathcal{N}$. The latter inequality ensures that the equilibrium transfer is nonnegative and is thus an element of $C(\mathcal{X}_n, \mathbb{R}_+)$, as required at the outset.

The recursive construction (6)–(7) has two difficulties. *First*, the resulting in-equilibrium transfer $\alpha = [\alpha_n^m]$ is potentially not Pareto optimal for the principals, in the sense that it may be possible to strictly lower in-equilibrium transfers to agents for some principals while making no other principal worse off. *Second*, the set of attainable in-equilibrium transfers using the recursive construction is in general strictly contained in the set $\mathcal{A}(\hat{x})$ of all in-equilibrium transfers that can be implemented as a WTE $(\hat{t}, \hat{x})$. Our direct equilibrium construction bypasses these shortcomings by providing a precise representation of *all* attainable WTEs of $\mathcal{G}$, as follows. To find all WTEs that implement an efficient outcome $\hat{x} \in \mathcal{X}$, it is by Proposition 1 necessary to find all excess transfers Δ whose elements Δ_n^m satisfy the $M+N$ inequalities (R). We thus provide a complete set of solutions to this "reduced contract design problem" (Section 3.4). In particular, we show that any solution to (R) can be represented as

[5]The reasoning behind expression (8) is that in equilibrium each principal m is minimizing her expenditure on transfers by paying only the amount needed to compensate agent n for the payoff difference, had he chosen his otherwise best action without principal m. The worst "punishment" principal m can inflict on agent n is thereby limited to paying him zero, since the transfer t_n^m is by assumption an element of $C(\mathcal{X}_n, \mathbb{R}_+)$.

an outcome-contingent convex combination of extremal basis functions, which implies a simple representation of all WTEs of $\mathcal{G}$, including an explicit representation of the set of attainable in-equilibrium transfers. In Section 3.5, we then provide a simple mapping from any solution Δ of (R) to any admissible equilibrium transfer $\hat{t} = \hat{\Delta} + \alpha$ with a Pareto-optimal in-equilibrium transfer α. Again, one finds that any equilibrium transfer in the subset of Pareto-optimal equilibrium transfers can be represented as a convex combination of extremal basis functions.

3.4 Solving the Reduced Contract Design Problem

We now construct a complete solution to the reduced contract design problem (R) given an efficient outcome $\hat{x}$. Any efficient outcome is thereby a maximizer of the joint surplus W on the compact set $\mathcal{X}$ of the agents' feasible actions. Since W is continuous, by Weierstrass Theorem (Berge 1963, p. 69) there exists at least one efficient outcome in $\mathcal{X}$. Fixing an efficient outcome $\hat{x}$, we look for solutions to the set of $M+N$ inequalities (R). For this it is useful to note that, by the efficiency of $\hat{x}$, any excess welfare $W(x) - W(\hat{x}) = \sum_{m \in \mathcal{M}} F^m(x) - \sum_{n \in \mathcal{N}} G_n(x_n)$ is nonpositive, i.e.,

$$\sum_{m \in \mathcal{M}} F^m(x) \leq \sum_{n \in \mathcal{N}} G_n(x_n) \tag{10}$$

for all $x \in \mathcal{X}$. Since our construction makes repeated use of the separating hyperplane theorem (Berge, 1963, p. 163), we assume that all gross payoff functions are concave (for a characterization of existence without such an assumption, see footnote 13).

ASSUMPTION 1 (PAYOFF CONCAVITY) *Principal m's gross payoff Π^m and agent n's gross payoff Γ_n are concave and bounded for all $(m,n) \in \mathcal{M} \times \mathcal{N}$.*

Naturally, Assumption 1 implies that the excess measures F^m and $-G_n$ in Proposition 1 are concave. A solution to (R) can now be obtained by induction on m. Since $F^1 \le \sum_{n\in\mathcal{N}} G_n - \sum_{m\ge 2} F^m$, F^1 is concave, $\sum_{n\in\mathcal{N}} G_n - \sum_{m\ge 2} F^m$ is convex, and because both functions vanish at $\hat{x}$, the separating hyperplane theorem implies the existence of vectors $\delta_n^1 \in \mathbb{R}^L$ for $n \in \mathcal{N}$ such that

$$F^1(x) \le \sum_{n\in\mathcal{N}} \langle \delta_n^1, x_n - \hat{x}_n\rangle \le \sum_{n\in\mathcal{N}} G_n(x_n) - \sum_{m\ge 2} F^m(x). \tag{11}$$

We set $\Delta_n^1(x_n) = \langle \delta_n^1, x_n - \hat{x}_n\rangle$ for $n \in \mathcal{N}$. In the second step of the induction, we observe that from (11),

$$F^2 \le \sum_{n\in\mathcal{N}} (G_n - \Delta_n^1) - \sum_{m\ge 3} F^m.$$

Moreover, F^2 is concave, whereas $\sum_{n\in\mathcal{N}}(G_n - \Delta_n^1) - \sum_{m\ge 3} F^m$ is convex (the functions Δ_n^1 are affine for all $n \in \mathcal{N}$), and both functions vanish at $\hat{x}$. Another application of the separating hyperplane theorem then implies the existence of vectors $\delta_n^2 \in \mathbb{R}^L$ for $n \in \mathcal{N}$ such that

$$F^2(x) \le \sum_{n\in\mathcal{N}} \langle \delta_n^2, x_n - \hat{x}_n\rangle \le \sum_{n\in\mathcal{N}} (G_n - \Delta_n^1)(x_n) - \sum_{m\ge 3} F^m(x). \tag{12}$$

We set $\Delta_n^2(x_n) = \langle \delta_n^2, x_n - \hat{x}_n\rangle$. Let us now describe a generic iteration of the induction. Suppose that for any $\mu \in \{1, \ldots, M-2\}$ we have affine functions Δ_n^m for $m \in \{1, \ldots, \mu - 1\}$ and $n \in \mathcal{N}$ such that

$$F^\mu \le \sum_{n\in\mathcal{N}} \left(G_n - \sum_{m\le \mu-1} \Delta_n^m \right) - \sum_{m\ge \mu+1} F^m.$$

By assumption, F^μ is concave and $\sum_{n\in\mathcal{N}}\left(G_n - \sum_{m\leq\mu-1}\Delta_n^m\right) - \sum_{m\geq\mu+1}F^m$ is convex. Both functions vanish at $\hat{x}$. The separating hyperplane theorem guarantees the existence of vectors $\delta_n^\mu \in \mathbb{R}^L$ for $n \in \mathcal{N}$ such that

$$F^\mu(x) \leq \sum_{n\in\mathcal{N}} \langle \delta_n^\mu, x_n - \hat{x}_n \rangle \leq \sum_{n\in\mathcal{N}} \left(G_n - \sum_{m\leq\mu-1} \Delta_n^m \right)(x_n) - \sum_{m\geq\mu+1} F^m(x).$$

This defines functions $\Delta_n^\mu(x_n) = \langle \delta_n^\mu, x_n - \hat{x}_n \rangle$ which are affine and vanish at $\hat{x}$, such that

$$F^\mu \leq \sum_{n\in\mathcal{N}} \Delta_n^\mu \leq \sum_{n\in\mathcal{N}} \left(G_n - \sum_{m\leq\mu-1} \Delta_n^m \right) - \sum_{m\geq\mu+1} F^m. \tag{13}$$

As a result,

$$F^{\mu+1} \leq \sum_{n\in\mathcal{N}} \left(G_n - \sum_{m\leq\mu} \Delta_n^m \right) - \sum_{m\geq\mu+2} F^m,$$

i.e., the induction hypothesis holds for $\mu+1$. We continue the induction until $\mu = M-1$ and obtain functions $\Delta_n^{M-1}(x_n) = \langle \delta_n^{M-1}, x_n - \hat{x}_n \rangle$ for $n \in \mathcal{N}$ that satisfy

$$F^{M-1} \leq \sum_{n\in\mathcal{N}} \langle \delta_n^{M-1}, x_n - \hat{x}_n \rangle \leq \sum_{n\in\mathcal{N}} \left(G_n - \sum_{m\leq M-2} \Delta_n^m \right) - F^M. \tag{14}$$

The algorithm terminates by setting

$$\Delta_n^M = G_n - \sum_{m\leq M-1} \Delta_n^m \tag{15}$$

for all $n \in \mathcal{N}$. We have therefore found MN functions Δ_n^m that vanish at $\hat{x}$. From relations (13)–(15), one obtains that

$$F^m \leq \sum_{n\in\mathcal{N}} \Delta_n^m$$

for all $m \in \mathcal{M}$. Moreover, equation (15) implies that

$$\sum_{m \in \mathcal{M}} \Delta_n^m = G_n$$

for all $n \in \mathcal{N}$. Consequently the excess transfer $\Delta = [\Delta_n^m]$ solves (R) and implements the efficient outcome $\hat{x}$, as intended. We have thus provided a constructive proof of the following result.

THEOREM 1 (EXISTENCE OF A WTE OF $\mathcal{G}$) . *Under Assumption 1, there exists a weakly truthful equilibrium* $(\hat{t}, \hat{x})$ *of the game* $\mathcal{G}$.

This construction is useful when the payoff functions are not differentiable at $\hat{x}$.[6] On the other hand, when the principals' payoffs are differentiable at the efficient outcome, it is possible to provide an explicit representation of an affine WTE.

ASSUMPTION 2 (PRINCIPAL PAYOFF REGULARITY) *Principal m's gross payoff* Π^m *is differentiable*[7] *at the efficient outcome* $\hat{x} \in \mathcal{X}$ *for all* $m \in \mathcal{M}$.

Since by the Rademacher theorem (Magaril-Il'yaev and Tikhomirov, 2003, p. 160) payoff concavity (i.e., Assumption 1) already implies differentiability of Π^m and Γ_n almost everywhere for all $(m, n) \in \mathcal{M} \times \mathcal{N}$, Assumption 2 is typically not a strong additional requirement. In case it is not satisfied, the following result can still be applied in a weaker form using subdifferentials.[8]

[6] This may for example occur if, at least for some m, the excess payoff F^m is the pointwise minimum of concave functions, two of which intersect at $\hat{x}$.

[7] If the efficient outcome under consideration is a noninterior point of the feasible set $\mathcal{X}$, then differentiability is to be interpreted with respect to any differentiable path of points in $\mathcal{X}$ leading to $\hat{x} \in \partial\mathcal{X}$. We implicitly assume that $\mathcal{X}$ is indeed path-connected in a neighborhood of $\hat{x}$.

[8] In that case one can assert that there exists an appropriate element of the subdifferential of F^m at $\hat{x}$ with respect to x_n, which can be used in (16) instead of the regular directional derivative.

Proposition 2 (Affine WTE of $\mathcal{G}$) *Under Assumptions 1–2, any efficient outcome $\hat{x} \in \mathcal{X}$ can be implemented as a weakly truthful equilibrium $(\hat{t}, \hat{x})$ of the game $\mathcal{G}$, where $\hat{t} = [\hat{t}_n^m]$ with $\hat{t}_n^m(x_n) = \hat{\Delta}_n^m(x_n) + \alpha_n^m$, where $\hat{\Delta}_n^m$ and α_n^m are given by the recursion (6)–(7) as a function of $\Delta = [\Delta_n^m]$, and*

$$\Delta_n^m(x_n) = \left\langle \frac{\partial F^m(\hat{x})}{\partial x_n}, x_n - \hat{x}_n \right\rangle \tag{16}$$

for all $(m, n) \in \mathcal{M} \times \mathcal{N}$.

Proof. By Assumption 2, the directional derivative $\partial F^m(\hat{x})/\partial x_n$ is well defined for all $(m, n) \in \mathcal{M} \times \mathcal{N}$. Moreover, the choice of $\delta_n^m = \partial F^m(\hat{x})/\partial x_n \in \mathbb{R}^L$ leads to excess transfers $\Delta_n^m(x_n) = \langle \delta_n^m, x_n - \hat{x}_n \rangle$ that satisfy

$$F^m(x) \leq \sum_{n \in \mathcal{N}} \Delta_n^m(x_n) \tag{17}$$

for all $m \in \mathcal{M}$ and all $x \in \mathcal{X}$. In fact, these excess transfers Δ_n^m are the *only* affine excess transfers vanishing at $\hat{x}$ that satisfy this inequality, for the lower epigraph of F^m is clearly supported at $\hat{x}$ by the hyperplane defined by the expression on the right-hand side of (17). Since $\sum_{m \in \mathcal{M}} F^m \leq \sum_{n \in \mathcal{N}} G_n$ by (10), the supporting hyperplane of $\sum_{m \in \mathcal{M}} F^m$, which is precisely $\sum_{(m,n) \in \mathcal{M} \times \mathcal{N}} \Delta_n^m$, lies below $\sum_{n \in \mathcal{N}} G_n$. That is,

$$\sum_{n \in \mathcal{N}} \left(\sum_{m \in \mathcal{M}} \Delta_n^m(x_n) \right) \leq \sum_{n \in \mathcal{N}} G_n(x_n) \tag{18}$$

for all $x \in \mathcal{X}$. Setting $x_j = \hat{x}_j$ for all $j \neq n$, (18) implies that $\sum_{m \in \mathcal{M}} \Delta_n^m(x_n) \leq G_n(x_n)$ for all $x_n \in \mathcal{X}_n$, since all the other terms vanish. This concludes our proof. ■

The affine solution (16) to the reduced contract design problem (R) is particularly simple and thus seems intuitively appealing. Nevertheless, since in-equilibrium transfers obtained from affine solutions of (R) are generally not Pareto-optimal, we have an interest in finding all other solutions to the reduced contract design problem. These additional solutions are either "below" the affine solution ("infra-affine solutions") or "above" the affine solution ("ultra-affine solutions"). "Mixed" solutions that are partially above and partially below the affine solution can of course be obtained by convex combination.[9]

Infra-Affine Excess Transfers. Reducing excess transfers relative to the affine solution can be accomplished by the following iterative process which, starting from any excess transfer that solves (R), yields a unique infra-affine lower limit $\underline{\Delta}$ that also solves (R). The construction provided here is somewhat related to the leveling algorithm of Diliberto and Straus (1951), which in the limit provides the best approximation of any function of two variables by a sum of two functions of one variable.[10] From a given excess transfer $\Delta = [\Delta_n^m]$, it is possible to obtain a new modified transfer $\tilde{\Delta} = [\tilde{\Delta}_n^m]$ with admissible out-of-equilibrium transfers below Δ as follows.

[9]Given two solution excess transfers Δ and $\check{\Delta}$, any convex combination $\lambda\Delta+(1-\lambda)\check{\Delta}$ for $\lambda \in (0,1)$ also constitutes a solution excess transfer.

[10]Kolmogorov (1957) (after some generalization by other authors) showed that it is possible to represent *any* continuous function f of n variables as a linear superposition of continuous functions of one variable and addition, i.e., $f(x_1, \ldots, x_n) = \sum_{k=1}^{2n+1} g\left(\sum_{l=1}^{n} \kappa_l \varphi_k(x_l)\right)$ for some appropriate continuous functions $\varphi_1, \ldots, \varphi_{2n+1}$, g, and constants $\kappa_1, \ldots, \kappa_n$. If addition is excluded (i.e., $g = 1$), as in our case, then one obtains a problem of best approximation via linear superposition which is difficult for $n \geq 3$ (Khavinson, 1997). Our problem is different from the standard formulation in that the approximation is constrained to be from above with a specified point of contact, for which to the best of our knowledge no prior results exist in the literature and our algorithm was obtained independently.

For each n, let $x_{-n} = (x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N)$, $\mathcal{X}_{-n} = \times_{j \neq n} \mathcal{X}_j$, and set

$$\tilde{\Delta}_n^m(x_n) = \max_{x_{-n} \in \mathcal{X}_{-n}} \left\{ F^m(x) - \sum_{j \neq n} \Delta_j^m(x_j) \right\} \tag{19}$$

for any single $(m, n) \in \mathcal{M} \times \mathcal{N}$. Then replace Δ_n^m in Δ by $\tilde{\Delta}_n^m$ and repeat (19) with the new Δ and another index $(m, n) \in \mathcal{M} \times \mathcal{N}$. From the definition of $\tilde{\Delta}_n^m$ the following result is immediate.

Lemma 1 (Excess-Transfer Decrement) *The modified excess transfer $[\tilde{\Delta}_n^m]$ is such that (i) $\tilde{\Delta}_n^m \leq \Delta_n^m$, and (ii) $\tilde{\Delta}_n^m + \sum_{j \neq n} \Delta_j^m \geq F^m$.*

Proof. (i) Since Δ solves (R), we obtain from (19) that

$$\tilde{\Delta}_n^m(x_n) \geq F^m(x) - \sum_{j \neq n} \Delta_j^m(x_j) \tag{20}$$

for all $x \in \mathcal{X}$. Moreover, since

$$\Delta_n^m(x_n) \geq F^m(x) - \sum_{j \neq n} \Delta_j^m(x_j)$$

for all x, taking the maximum on the right-hand side with respect to x_{-n}, shows that $\Delta_n^m(x_n) \geq \tilde{\Delta}_n^m(x_n)$ for all x_n. (ii) This assertion follows directly from (20). ■

We denote by $\tilde{\Delta}$ the matrix obtained after successive application of (19) for each $(m, n) \in \mathcal{M} \times \mathcal{N}$. It is clear that starting from any infra-affine solution $\Delta = [\Delta_n^m]$, the new excess transfer $\tilde{\Delta} = [\tilde{\Delta}_n^m]$ remains an infra-affine solution to (R). We also note that the monotonicity in part (i) of Lemma 1 is pointwise. Replacing Δ_n^m by $\tilde{\Delta}_n^m$, we can repeat this procedure with another index $j \neq n$, which leads to the following sequential algorithm:

1. Set $m = n = 1$.

2. Compute $\tilde{\Delta}_n^m$.

3. Replace Δ_n^m by the function found in 2.

4. If $n < N$, increase n by one, or else if $m < M$ set $n = 1$ and increase m by 1, otherwise terminate.

5. Go back to step 2 with the new values of m and n.

This sequential algorithm, which has MN steps, can be itself iterated, yielding a sequence of excess transfer matrices $\sigma(\Delta) = \{\tilde{\Delta}_{(k)}(\Delta)\}_{k=0}^{\infty}$ with $\tilde{\Delta}_0 = \Delta$, where the starting matrix Δ satisfies the system of inequalities (R). The following result asserts that the limit $\Delta_\infty(\Delta) = \lim_{k\to\infty} \Delta_{(k)}(\Delta)$ is well defined and constitutes an admissible (i.e., continuous) equilibrium excess transfer.

Proposition 3 (Lower Excess-Transfer Bound) *For any admissible equilibrium excess transfer Δ, the limit $\Delta_\infty(\Delta)$ of the sequence $\sigma(\Delta)$ exists, is continuous on $\mathcal{X}$, and solves the reduced contract design problem (R).*

Proof. Let Δ be any admissible equilibrium excess transfer. For each (n, m), the corresponding component of the sequence $\sigma(\Delta)$ is bounded from below

by

$$F^m(\hat{x}_1, \ldots, \hat{x}_{n-1}, x_n, \hat{x}_{x+1}, \ldots, \hat{x}_N) - \sum_{j \neq n} \Delta(\hat{x}_j).$$

Moreover, the sequence is nonincreasing pointwise by Lemma 1. Therefore, there exists a function $\Delta_\infty = \lim_{k\to\infty} \Delta_{(k)}$, defined as the pointwise limit of $\sigma(\Delta)$. By (19) the limit function $\Delta_\infty = [\Delta^m_{n,\infty}]$ satisfies

$$\Delta^m_{n,\infty}(x_n) = \sup_{x_{-n}\in\mathcal{X}_{-n}} \left\{ F^m(x) - \sum_{j\neq n} \Delta^m_{j,\infty}(x_j) \right\} \tag{21}$$

for all $(m,n) \in \mathcal{M} \times \mathcal{N}$. Since F^m is continuous on $\mathcal{X}$ by hypothesis, for any $x \in \mathcal{X}$ and any $\varepsilon > 0$ there exists a real number $\rho(x,\varepsilon) > 0$ such that

$$\|x - \bar{x}\| \leq \rho(x,\varepsilon) \quad \Rightarrow \quad |F^m(x) - F^m(\bar{x})| \leq \varepsilon \tag{22}$$

for all $x, \bar{x} \in \mathcal{X}$. Thus, as long as $\|x-\bar{x}\| = \|(x_n, x_{-n})-(\bar{x}_n, \bar{x}_{-n})\| \leq \rho((x_n, y(x_n)), \varepsilon)$,

$$\Delta^m_{n,\infty}(x_n) - \Delta^m_{n,\infty}(\bar{x}_n) \leq F^m(x_n, y(x_n)) - F^m(\bar{x}_n, y(x_n)) \leq \varepsilon. \tag{23}$$

Similarly, $\|x - \bar{x}\| = \|(x_n, x_{-n}) - (\bar{x}_n, \bar{x}_{-n})\| \leq \rho((\bar{x}_n, y(\bar{x}_n)), \varepsilon)$ implies that

$$\Delta^m_{n,\infty}(\bar{x}_n) - \Delta^m_{n,\infty}(x_n) \leq F^m(\bar{x}_n, y(\bar{x}_n)) - F^m(x_n, y(\bar{x}_n)) \leq \varepsilon, \tag{24}$$

where

$$y(x_n) \in \arg\sup_{x_{-n}\in\mathcal{X}_{-n}} \left\{ F^m(x_n, x_{-n}) - \sum_{j\neq n} \Delta^m_{j,\infty}(x_j) \right\}.$$

Since the set $\mathcal{X}$ is compact by assumption, $y(x_n)$ exists and lies in $\mathcal{X}_{-n}$ for any $x_n \in \mathcal{X}_n$. Moreover, compactness of $\mathcal{X}$ and (22) imply that F^m is uniformly continuous on $\mathcal{X}$, simply replacing $\rho(x,\varepsilon)$ by $\bar{\rho}(\varepsilon) = \inf_{x\in\mathcal{X}} \rho(x,\varepsilon) > 0$. Hence, using relations (23) and (24), we obtain that for any $x, \bar{x} \in \mathcal{X}$,

$$\|x - \bar{x}\| \leq \bar{\rho}(\varepsilon) \quad \Rightarrow \quad |\Delta^m_{n,\infty}(x_n) - \Delta^m_{n,\infty}(\bar{x}_n)| \leq \varepsilon,$$

so that the excess transfer $\Delta_\infty = [\Delta^m_{n,\infty}]$ is uniformly continuous on $\mathcal{X}$. It also solves (R), since each excess transfer $\Delta_{(k)}$ solves (R) by construction. In other words, Δ_∞ is an admissible equilibrium excess transfer. Via the recursion (6)–(7), Δ_∞ defines a WTE implementing the efficient outcome $\hat{x}$ that was used to generate the starting matrix Δ. This completes our proof. From (21), it is easy to show that any limit obtained by the above procedure is *minimal*, in the sense that if some functions $[\phi^m_n]$, $\phi^m_n : \mathcal{X}_n \to \mathbb{R}$, satisfy

$$\sum_{n \in \mathcal{N}} \phi^m_n \geq F^m \tag{25}$$

for any $m \in \mathcal{M}$, and if

$$\phi^m_n \leq \Delta^m_n \tag{26}$$

for any $(m, n) \in \mathcal{M} \times \mathcal{N}$, then

$$\phi^m_n = \Delta^m_n.$$

Now consider two transfers $[\hat{\Delta}^m_n]$ and $[\check{\Delta}^m_n]$ that are both the limit of some sequence described in (i). Then, the transfer $[\underline{\Delta}^m_n]$ defined by $\underline{\Delta}^m_n = \min\{\hat{\Delta}^m_n, \check{\Delta}^m_n\}$ satisfies (25) and (26) for any $[\phi^m_n] \in \{[\hat{\Delta}^m_n], [\check{\Delta}^m_n]\}$, which implies that $\underline{\Delta}^m_n = \hat{\Delta}^m_n = \check{\Delta}^m_n$ for all $(m, n) \in \mathcal{M} \times \mathcal{N}$. ■

In general, the limit $\Delta_\infty(\Delta)$ is not unique: it depends on the starting point Δ. In order to simplify the analysis for the description of the WTE implementing an efficient outcome, we restrict our attention to the limit $\underline{\Delta} = \Delta_\infty(\Delta_{\text{aff}})$ where Δ_{aff} is the matrix of affine excess transfers obtained by Proposition 2. Under some conditions on the payoff functions given in Assumption 3, the lower limit $\Delta_\infty(\Delta)$ is in fact independent of Δ, in which case the restriction is vacuous. For example, $\Delta_\infty(\Delta)$ is clearly independent of Δ when the payoff functions $\{F^m\}_{m \in \mathcal{M}}$ are additive in $(x_1, \ldots, x_N)$.

However, the following proposition shows that independence holds under more general conditions.

ASSUMPTION 3 (PAYOFF SPECIFICATION) *Each principal m's excess payoff function $F^m : \mathcal{X} \to \mathbb{R}$ is of the form*

$$F^m(x_1, \ldots, x_N) = f^m\left(g_1^m(x_1), \ldots, g_N^m(x_N)\right),$$

where f^m is submodular[11] *and vanishes at $(0, \ldots, 0)$, and the function $g_n^m : \mathcal{X}_n \to \mathbb{R}_+$ vanishes at $\hat{x}_n$ for all $n \in \mathcal{N}$.*

PROPOSITION 4 *Suppose that Assumption 3 holds. Then, $\Delta_\infty(\Delta)$ is independent of Δ.*

Proof. From Proposition 8, the "additive upper envelope" $\sum_n \underline{\Delta}_n^m$ of each F^m is unique. Letting $\underline{\Delta} = [\underline{\Delta}_n^m]$, $\underline{\Delta}$ is clearly below any limit $\Delta_\infty(\Delta)$. Since $\Delta_\infty(\Delta)$ is minimal, it must equal $\underline{\Delta}$. ■

Section 3.6 provides several examples where Assumption 3 holds. For the remainder of the chapter, we will only consider the lower excess-transfer bound $\underline{\Delta}$ generated by the affine excess transfer of Proposition 2. By Proposition 3, $\underline{\Delta}$ is infra-affine. We thus obtain a general class of WTEs implementing the efficient outcome. When Assumption 3 holds, this set characterizes the set of *all* WTEs implementing the efficient outcome.[12] Note also that, by construction, $\underline{\Delta}$ solves the reduced contract

[11] A function h of k variables $(x_1, \ldots x_k) \in \mathbb{R}^{qk}$, where q is a positive integer, is *submodular* if and only if for any $(x, y) \in \mathbb{R}^{2qk}$, $h(\min(x, y)) + h(\max(x, y) \leq h(x) + h(y)$, where the minimum and maximum are taken componentwise.

[12] In principle, it is also possible to consider the distinct lower limits in order to describe all WTEs, but this task is likely to involve a continuum of lower limit functions, depending on the form of the payoff functions.

design problem (R). When the limit obtained by the above procedure is independent of Δ, that limit is the smallest additive function lying above φ. For that reason, we call this limit the *additive upper envelope* of φ. Section 3.6 provides a way to directly compute the additive upper envelope when φ satisfies[13] Assumption 3.

Ultra-Affine Excess Transfers. The solution constructed in the proof of Theorem 1 was affine for all principals except possibly the last. By relabelling principals such that the role of the last is assigned to principal $i \in \mathcal{M}$, we obtain in general M different solutions, which all lie above the affine solution described in Proposition 2. An outcome-contingent convex combination of these solutions is given by

$$\Delta_n^m(x_n;\theta_n^m) = \langle \frac{\partial F^m(\hat{x})}{\partial x_n}, x_n - \hat{x}_n \rangle + \theta_n^m(x_n)\left(G_n(x_n) - \sum_{i\in\mathcal{M}} \langle \frac{\partial F^i(\hat{x})}{\partial x_n}, x_n - \hat{x}_n \rangle\right) \quad (27)$$

for all $(m,n) \in \mathcal{M} \times \mathcal{N}$, whereby $\theta_n^m \in C(\mathcal{X}_n, [0,1])$ such that $\sum_{m\in\mathcal{M}} \theta_n^m = 1$. It is clear that each $\Delta_n^m(\cdot;\theta_n^m)$ is ultra-affine, as a convex combination of excess transfers that are ultra-affine by construction. Replacing the affine excess transfers of the form (16) in (27) by the infra-affine lower limit $\underline{\Delta}_n^m$, we obtain the tight *upper bound* for solutions to the reduced contract design problem

$$\bar{\Delta}_n^m(x_n;\theta_n^m) = \underline{\Delta}_n^m(x_n) + \theta_n^m(x_n)\left(G_n(x_n) - \sum_{i\in\mathcal{M}} \underline{\Delta}_n^i(x_n)\right) \quad (28)$$

for all $x_n \in \mathcal{X}_n$.

[13] The additive upper envelope $\underline{\Delta}$ is very much related to the existence of a WTE in the absence of concavity (Assumption 1), provided the payoff functions are at least continuously differentiable. Indeed, it is clear from our construction of the lower limit (which did not depend on payoff concavity) that a WTE of $\mathcal{G}$ exists if $\underline{\Delta} = \Delta_\infty(\Delta)$ solves (R), where the starting matrix $\Delta = [\Delta_n^m]$ can be any excess transfer matrix that satisfies the first inequality in (R). For instance, one could take $\Delta_n^m(x_n) = k\|x_n - \hat{x}_n\|$ for a large enough positive constant k, since all excess payoffs are Lipschitz on the compact set of outcomes. Moreover, if a WTE exists, then there should be a lower limit that solves (R). In particular, under Assumption 3, the fact that $\underline{\Delta}$ solves (R) is also a necessary condition.

PROPOSITION 5 (UPPER EXCESS-TRANSFER BOUND) *The excess transfer* $[\tilde{\Delta}_n^m(\cdot;\theta_n^m)]$ *solves the reduced contract design problem (R) for any matrix of outcome-contingent weights* $\theta_n^m \in C(\mathcal{X}_n,[0,1])$ *with* $\sum_{m\in\mathcal{M}}\theta_n^m = 1$.

Proof. We prove more generally that, given *any* solution $\Delta = [\Delta_n^m]$ of (R), the excess transfer $\tilde{\Delta} = [\tilde{\Delta}_n^m]$ with

$$\tilde{\Delta}_n^m = \Delta_n^m + \theta_n^m \left(G_n - \sum_{i\in\mathcal{M}} \Delta_n^i\right) \tag{29}$$

solves (R) provided that the weights $\theta_n^m \in C(\mathcal{X}_n,[0,1])$ satisfy $\sum_{m\in\mathcal{M}}\theta_n^m = 1$. Our assertion then follows immediately for $\Delta = \underline{\Delta}$. Since $G_n \geq \sum_{i\in\mathcal{M}}\Delta_n^m$ by (R), $\tilde{\Delta}_n^m \geq \Delta_n^m$. Thus,

$$\sum_{n\in\mathcal{N}} \tilde{\Delta}_n^m \geq \sum_{n\in\mathcal{N}} \Delta_n^m \geq F^m,$$

for all $m \in \mathcal{M}$, since Δ satisfies (R). Moreover, for any $n \in \mathcal{N}$ the identity

$$\sum_{m\in\mathcal{M}} \tilde{\Delta}_n^m = G_n,$$

holds for any admissible θ, so that the excess transfer $\tilde{\Delta}$ solves the reduced contract design problem (R). ■

The lower and upper bounds for solutions to the reduced contract design problem (R) can now be used to construct an *extremal basis* $\mathcal{B}$ containing $M+1$ transfers, as follows.

Let

$$\mathcal{B} = \{[B_n^{m,0}], \ldots, [B_n^{m,M}]\}, \tag{30}$$

where

$$B_n^{m,\mu}(\cdot) = \begin{cases} \bar{\Delta}_n^m(\cdot;1), & \text{if } m = \mu, \\ \bar{\Delta}_n^m(\cdot;0), & \text{otherwise,} \end{cases} \tag{31}$$

for all $(m, \mu, n) \in \mathcal{M} \times \bar{\mathcal{M}} \times \mathcal{N}$ with the abbreviation $\bar{\mathcal{M}} = \{0\} \cup \mathcal{M}$. The central result of this section is that any convex combination of elements of $\mathcal{B}$ is a solution to (R), and that the converse is true under Assumption 3. For each n, let $\hat{\mathcal{X}}_n = \{x_n \in \mathcal{X}_n : \sum_{m\in\mathcal{M}} \underline{\Delta}_n^m(x_n) < G_n(x_n)\}$.

Definition 1 (Admissible Weights) *A matrix $\theta = [\theta_n^m]$ of functions defines admissible weights, denoted $\theta \in \mathcal{W}$, if its entries satisfy*

(i) for each (m, n) in $\mathcal{M} \times \mathcal{N}$, $\theta_n^m : \mathcal{X}_n \to [0, 1]$;

(ii) for all n, $\sum_{m\in\mathcal{M}} \theta_n^m = 1$;

(iii) for each (m, n) in $\mathcal{M} \times \mathcal{N}$, θ_n^m is continuous on $\hat{\mathcal{X}}_n$.

Theorem 2 (Representation of Reduced Equilibrium Contracts) *Suppose that Assumption 1 holds, and let*

$$\operatorname{co} \mathcal{B} = \left\{ [\Delta_n^m] : \Delta_n^m = \sum_{\mu\in\bar{\mathcal{M}}} \vartheta_n^\mu B_n^{m,\mu}, [\vartheta_n^\mu] \in \mathcal{W} \right\}. \tag{32}$$

Then: (i) any element of $\operatorname{co}\mathcal{B}$ solves the reduced contract design problem (R), and (ii) when the lower limit $\Delta_\infty(\Delta)$ is independent of Δ, the set of solutions to the reduced contract design problem (R) is exactly $\operatorname{co}\mathcal{B}$.

Proof. We start by proving (ii): Let $[\Delta_n^m]$ be a solution to the reduced contract design problem (R). We need to show that $[\Delta_n^m] \in \operatorname{co}\mathcal{B}$. Indeed, for any fixed $n \in \mathcal{N}$ and $x_n \in \mathcal{X}_n$, let $\beta = (\Delta_n^1(x_n), \dots, \Delta_n^M(x_n))$ and $\beta^\mu = (B_n^{1,\mu}(x_n), \dots, B_n^{M,\mu}(x_n))$ for all $\mu \in \bar{\mathcal{M}}$. These $M+2$ vectors all belong to $\mathbb{R}^M$. Since $\beta^0 \leq \beta$ and $\langle \beta, (1, \dots, 1) \rangle = \sum_{m \in \mathcal{M}} \Delta_n^m(x_n) \leq G_n(x_n)$, the vector β belongs to the M-simplex with vertices $\beta^0, \dots, \beta^M$. As a result, there exist $M+1$ nonnegative numbers $\vartheta_n^0(x_n), \dots, \vartheta_n^M(x_n)$ with $\sum_{\mu \in \bar{\mathcal{M}}} \vartheta_n^\mu(x_n) = 1$, such that $\beta = \sum_{\mu \in \bar{\mathcal{M}}} \vartheta_n^\mu(x_n) \beta^\mu$. If $x_n \in \hat{\mathcal{X}}_n$, then the simplex is nondegenerate in a neighborhood of x_n, which implies uniqueness and continuity of the coefficient functions $\vartheta_n^0, \dots, \vartheta_n^M$ in the representation (32) on that neighborhood.[14]

We now prove (i): Consider any element $\Delta \in \operatorname{co}\mathcal{B}$. Using (30)–(32),

$$\Delta_n^m(x_n) = \sum_{\mu \in \bar{\mathcal{M}}} \vartheta_n^\mu(x_n) B_n^{m,\mu}(x_n) = \vartheta_n^m(x_n) \bar{\Delta}_n^m(x_n; 1) + \bar{\Delta}(x_n; 0)(1 - \vartheta_n^m(x_n))$$

on $\mathcal{X}_n$, and thus

$$\Delta_n^m = \vartheta_n^m \left(G_n - \sum_{i \neq m} \underline{\Delta}_n^i \right) + (1 - \vartheta_n^m) \underline{\Delta}_n^m = \underline{\Delta}_n^m + \vartheta_n^m \left(G_n - \sum_{i \in \mathcal{M}} \underline{\Delta}_n^i \right) \tag{33}$$

for all $(m, n) \in \mathcal{M} \times \mathcal{N}$, where $\sum_{m \in \mathcal{M}} \vartheta_n^m \leq 1$. Since by Proposition 3 the transfer $\underline{\Delta}$ solves the reduced contract design problem, (33) implies that $\Delta_n^m \geq \underline{\Delta}_n^m$, whence

$$\sum_{n \in \mathcal{N}} \Delta_n^m \geq \sum_{n \in \mathcal{N}} \underline{\Delta}_n^m \geq F^m$$

[14]If $F = \sum_{m \in \mathcal{M}} F^m$ is strictly concave in a neighborhood of the efficient outcome $\hat{x}$, it is easily shown that the simplex is everywhere nondegenerate.

for all $n \in \mathcal{N}$. In addition,

$$\sum_{m \in \mathcal{M}} \Delta_n^m = (1 - \vartheta_n^0) G_n + \vartheta_n^0 \sum_{i \in \mathcal{M}} \underline{\Delta}_n^i = G_n - \vartheta_n^0 \left(G_n - \sum_{i \in \mathcal{M}} \underline{\Delta}_n^i \right) \leq G_n,$$

where $\vartheta_n^0 = 1 - \sum_{m \in \mathcal{M}} \vartheta_n^m$ takes values in $[0,1]$. Thus, $\Delta = [\Delta_n^m]$ indeed solves the reduced contract design problem (R). The continuity of Δ_n^m comes from the continuity of the weights on $\hat{\mathcal{X}}_n$, and the fact that all convex combination yield the same point when the simplex is degenerate (i.e., $x_n \in \mathcal{X}_n \setminus \hat{\mathcal{X}}_n$). ■

Theorem 2 provides, when Assumption 3 holds, a simple characterization of all solutions to the reduced contract design problem (R). We denote $\mathcal{R}'$ the subset of excess transfers solving (R) which have the representation (32). Which particular equilibrium excess transfer Δ to choose depends on the desired allocation of surplus at the implemented efficient equilibrium outcome $\hat{x}$. The next section constructs Pareto optimal WTEs based on the obtained reduced transfers.

From Theorem 2, any element $\Delta = [\Delta_n^m]$ of $\mathcal{R}'$ can be represented as

3.5 Constructing Weakly Truthful Equilibria

$$\Delta_n^m = \underline{\Delta}_n^m + \vartheta_n^m \left(G_n - \sum_{i \in \mathcal{M}} \underline{\Delta}_n^i \right),$$

where $[\vartheta_n^m]$ defines admissible weights (over $\bar{\mathcal{M}}$ instead of $\mathcal{M}$). From (8), the in-equilibrium transfer α_n^m for any solution to $\mathcal{R}'$ is given by

$$\alpha_n^m = -\min_{x_n \in \mathcal{X}_n} \left\{ \underline{\Delta}_n^m + (\vartheta_n^m + \vartheta_n^0) \left(G_n - \sum_{i \in \mathcal{M}} \underline{\Delta}_n^i \right) \right\}, \tag{34}$$

where $\vartheta_n^0 = 1 - \sum_{i \in \mathcal{M}} \vartheta_n^i$, as in the previous section. For any given agent n, principal m prefers the smallest possible in-equilibrium transfer α_n^m. Therefore, by reducing ϑ_n^0, it is possible to simultaneously reduce in-equilibrium payments for all principals, which thus leads to (weak) Pareto-improvements. To find Pareto-optimal WTEs of $\mathcal{R}'$, we can thus restrict our attention to the frontier $\mathcal{F} = \{[\Delta_n^m] \in \operatorname{co} \mathcal{B} : [\vartheta_n^0] = 0\}$. In general, however, $\mathcal{F}$ contains elements of $\mathcal{R}'$ which are not Pareto optimal. Nevertheless, using the following procedure it is possible to filter out all Pareto-optimal allocations. For the rest of this section, we construct the set of WTEs in $\mathcal{R}'$ which are Pareto optimal. For this, we define

$$L_n^m(\Delta) = \min_{x_n \in \mathcal{X}_n} \Delta_n^m(x_n). \tag{35}$$

For $[\Delta_n^m] \in \mathcal{F}$ fixed, we observe from (8) that $\alpha_n^m = -\min_{x_n \in \mathcal{X}_n} \Delta_n^m(x_n)$. For each $(m, n) \in \mathcal{M} \times \mathcal{N}$, define $D = [D_n^m]$ by

$$D_n^m(x_n) = \max\{\underline{\Delta}_n^m(x_n), L_n^m(\Delta)\}. \tag{36}$$

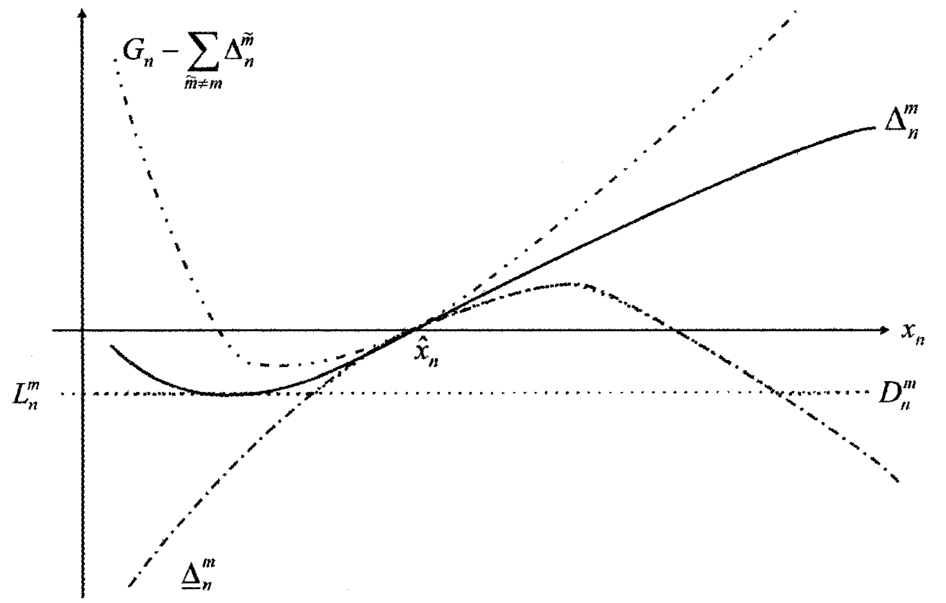

Figure 1: Ironing Algorithm.

By construction, $D_n^m \leq \Delta_n^m$ but $L_n^m(D) = L_n^m(\Delta)$. The previous steps are illustrated by Figure 1. A decrease in any D_n^m causes an increase in α_n^m, making principal m strictly worse off (when changes are limited to transactions with agent n). However, $[D_n^m]$ does not necessarily lie on the frontier $\mathcal{F}$. Therefore, it might be possible to Pareto-improve on $[D_n^m]$. For $(m,n) \in \mathcal{M} \times \mathcal{N}$, let $\mathcal{X}_n^m = \{x_n \in \mathcal{X}_n : D_n^m(x_n) = L_n^m(D)\}$, let[15]

$$d_n^m(D) = \min_{x_n \in \mathcal{X}_n^m} \left\{ G_n(x_n) - \sum_{\mathcal{M}} D_n^m(x_n) \right\} \geq 0, \tag{37}$$

and denote by $\underline{x}_n^m$ a point that reaches this minimum. If $d_n^m(D) = 0$, then L_n^m cannot be increased without making another principal worse off, since at $\underline{x}_n^m$, $\sum_{m \in \mathcal{M}} D_n^m = G_n$ and $D_n^m = L_n^m$. If, however, $d_n^m(D) > 0$, then it is possible to increase L_n^m by setting

$$\tilde{D}_n^m = \max\{L_n^m(D) + d_n^m(D), \underline{\Delta}_n^m\}.$$

[15]The dependence on m comes from the maximization domain $\mathcal{X}_n^m$.

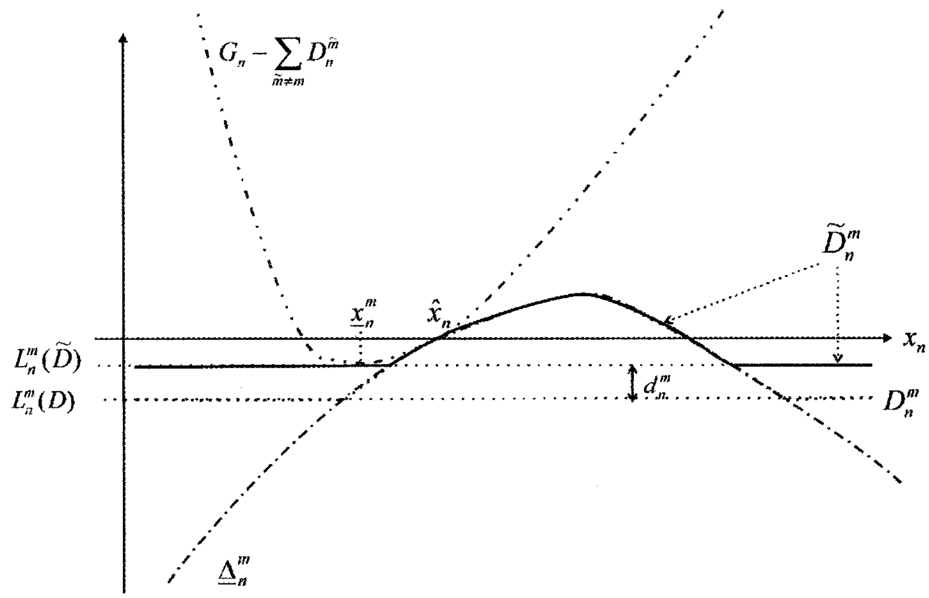

Figure 2: Constructing Pareto-optimal transfers.

Figure 2 illustrates the previous steps. Performing this transformation with $m = 1$, we then replace D_n^1 by $\tilde{D}_n^1$ and repeat this step for $m = 2, \ldots, M$. This sequential procedure yields new transfers $\tilde{D}_n^m$ for fixed n, and in-equilibrium transfers that are Pareto optimal in $\mathcal{R}'$ when attention is restricted to agent n.

LEMMA 2 (n-PARETO OPTIMALITY) *If*

$$L_n^m(\hat{D}) \geq L_n^m(\tilde{D}) \quad \forall m \in \mathcal{M}$$

for some excess transfers $(\hat{D}_n^m)_{m \in \mathcal{M}}$*, then*

$$L_n^m(\hat{D}) = L_n^m(\tilde{D}) \quad \forall m \in \mathcal{M}.$$

Proof. We first observe that each step weakly increases the transfers, so that d_n^i remains at zero for $i < m$ when the m^{th} step occurs. By construction, $d_n^m(\tilde{D}) = 0$ for all m, so that $L_n^m(\tilde{D})$ cannot be increased without making another principal worse off. ■

In order to stay in $\mathcal{F}$, we conclude the procedure by letting $\tilde{D}_n^M = G_n - \sum_{m \neq M} \tilde{D}_n^m$. Note that this last transformation does not affect in-equilibrium transfers since $\tilde{D}_n^m$ was already n-Pareto optimal, and increasing $\tilde{D}_n^M$ is a weak n-Pareto improvement. We can perform the procedure for all $n \in \mathcal{N}$ and obtain a new excess transfer matrix $[\tilde{D}_n^m]$. Denote by $\tilde{T} : \mathcal{F} \to \mathcal{F}$ the operator that maps $\Delta = [\Delta_n^m]$ to $\tilde{D} = [\tilde{D}_n^m]$. To state our result in its more general form, we define equivalence classes of excess transfer matrices.

Definition 2 (Excess-Transfer Equivalence) *Two excess transfer matrices* $[\Delta_n^m]$ *and* $[\tilde{\Delta}_n^m]$ *in* $\mathcal{F}$ *are equivalent, denoted* $[\Delta_n^m] \equiv [\tilde{\Delta}_n^m]$, *if* $\alpha_n^m(\Delta) = \alpha_n^m(\tilde{\Delta})$ *for all* $(m, n) \in \mathcal{M} \times \mathcal{N}$.

Note that an equivalence class is entirely characterized by the $M \times N$-dimensional matrix $\alpha = [\alpha_n^m]$ of in-equilibrium transfers. Let

$$\mathcal{A}_n = \{(\alpha_n^1(\Delta), \ldots, \alpha_n^M(\Delta)) : \Delta \in \mathcal{R}'\}$$

be the set of in-equilibrium transfers to agent n that are reachable from $\mathcal{R}'$, and let $\mathcal{A} = \prod_1^N \mathcal{A}_n$ be the cartesian product of these sets.

DEFINITION 3 *An equivalence class* α *of excess transfers is agent-wise Pareto optimal (in* $\mathcal{R}'$*), denoted* $\alpha \in \mathcal{L}$*, if for each* n *the vector* $\alpha_n = (\alpha_n^1, \ldots, \alpha_n^M)$ *belongs to the lower boundary* $\partial_- \mathcal{A}_n \subset \mathbb{R}_+^{ML}$ *of* $\mathcal{A}_n$*, defined by*

$$\partial_- \mathcal{A}_n = \{\alpha_n \in \mathcal{A}_n : \nexists \tilde{\alpha}_n \in \mathcal{A}_n, \tilde{\alpha}_n < \alpha_n\}.$$

Denoting $T : \mathcal{F} \to \mathcal{A}$ the operator that maps $\Delta = [\Delta_n^m]$ to $[\alpha_n^m(\tilde{T}(\Delta))]$, we have the following result.

THEOREM 3 (PARETO-OPTIMAL IN-EQUILIBRIUM TRANSFERS) *The image* $T(\mathcal{F})$ *is exactly the set of agent-wise Pareto-optimal in-equilibrium transfers.*

Proof. From Lemma 2, the image of T is clearly included in $\mathcal{L}$. Now take any in-equilibrium transfer matrix α in $\mathcal{L}$ and let Δ be an excess transfer matrix that implements it. By construction, $\tilde{T}(\Delta)$ is a weak Pareto improvement of Δ. Since α is agent-wise Pareto optimal, $T(\Delta)$ is less than α componentwise, and thus $T(\Delta) = \alpha$. ∎

If each principal m only cares about the sum $\alpha_1^m + \cdots + \alpha_N^m$ of her in-equilibrium transfers, then agent-wise Pareto optimality does *not* imply Pareto optimality. Once the lower boundaries $\partial_- \mathcal{A}_n$ of each $\mathcal{A}_n$ have been constructed, the set of Pareto-optimal excess transfer matrices corresponds to the lower boundary of the sum $\partial_- \mathcal{A}_1 + \cdots + \partial_- \mathcal{A}_N \in \mathbb{R}_+^{ML}$. We now determine the "absolute" lower boundaries in $\mathcal{R}'$ for each principal. This is the best scenario for a given principal, where only her profit is taken into account to implement the efficient outcome.

Proposition 6 (Transfer Lower Bound) *For each $n \in \mathcal{N}$, the lowest in-equilibrium transfer that can be obtained by principal m is*

$$\underline{\alpha}_n^m = -\min_{x_n \in \mathcal{X}_n} \left\{ G_n(x_n) - \sum_{i \neq m} \underline{\Delta}_n^i(x_n) \right\}.$$

Proof. $G_n(x_n) - \sum_{i \neq m} \underline{\Delta}_n^i(x_n)$ is the highest excess transfer $\bar{\Delta}_n^m$ that can be attained by principal m, and thus yields the lowest α_n^m. Moreover, $\alpha_n^m = -\min\left\{\bar{\Delta}_n^m\right\}$ since, in that case, the sum of the excess transfers is equal to G_n. ■

We can also derive an upper bound for $\partial_- \mathcal{A}_n$, implying a limit on the worst case for principal m in an n-Pareto-optimal transfer.

Proposition 7 (Transfer Upper Bound) *For each $n \in \mathcal{N}$, $\partial_- \mathcal{A}_n$ is bounded from above by $\bar{\alpha}_n = (\bar{\alpha}_n^1, \ldots, \bar{\alpha}_n^M)$, where*

$$\bar{\alpha}_n^m = -\min_{x_n \in \mathcal{X}_n} \underline{\Delta}_n^m(x_n).$$

Proof. For any excess transfer Δ in $\mathcal{F}$,

$$\alpha_n^m(\Delta) = -\min_{x_n \in \mathcal{X}_n} \Delta_n^m(x_n) \leq -\min_{x_n \in \mathcal{X}_n} \underline{\Delta}_n^m(x_n) = \bar{\alpha}_n^m.$$

Since $\tilde{T}(\Delta) \geq \Delta$,

$$\alpha_n^m\left(\tilde{T}(\Delta)\right) \leq \alpha_n^m(\Delta).$$

Therefore, $T(\Delta) \leq \bar{\alpha}_n$ for all $\Delta \in \mathcal{F}$. By Theorem 3, the set of agent-wise Pareto-optimal transfers is precisely the image of T, which concludes our proof. ■

In the next result, we construct an excess transfer matrix Δ that attains a given in-equilibrium transfer α in $T(\mathcal{F})$. Since the construction can be done agent-wise, we fix $n \in \mathcal{N}$ and construct $(\Delta_n^m)_{m\in\mathcal{M}}$. Thus suppose that $\alpha_n = (\alpha_n^m)_{m\in\mathcal{M}}$ is a given vector in the boundary $\partial_{-}\mathcal{A}_n$. Let

$$\Delta_n^m(x_n) = \max\{\underline{\Delta}_n^m(x_n), -\alpha_n^m\} \tag{38}$$

for all $x_n \in \mathcal{X}_n$. By construction, we have

$$-\min_{x_n\in\mathcal{X}_n} \Delta_n^m(x_n) \geq \alpha_n^m. \tag{39}$$

Moreover,

$$\underline{\Delta}_n^m \leq \Delta_n^m, \tag{40}$$

and Δ_n^m is the smallest excess transfer that satisfies (39) and (40). Since by assumption α_n is attainable, this implies that $\sum_{m\in\mathcal{M}} \Delta_n^m \leq G_n$. To ensure that Δ_n^m belongs to $\mathcal{F}$, we redefine Δ_n^M as

$$\Delta_n^M = G_n - \sum_{m\neq M} \Delta_n^m. \tag{41}$$

Moreover, n-Pareto optimality of α_n implies that in fact equality obtains in (39). Therefore, $(\Delta_n^m)_{m\in\mathcal{M}}$ implements α_n. From the construction, all the excess transfers are clearly continuous, which together with (R) implies their admissibility. We have thus shown the following result.

Theorem 4 (Pareto-Optimal Equilibrium Implementation) *Suppose that α_n belongs to the set $\partial_{-}\mathcal{A}_n$ of n-Pareto optimal in-equilibrium transfers. Then, the vector of excess transfers $(\Delta_n^m)_{m\in\mathcal{M}}$ defined by (38) for $m < M$ and by (41) for $m = M$ results in a WTE with in-equilibrium transfer α_n.*

Theorem 4 provides a practical implementation of Pareto-optimal equilibrium contracts for the principals.

3.6 Exact Lower Excess-Transfer Bound

The exact representation of all WTEs in Sections 3.4 and 3.5 relies on the existence of a *unique* lower excess-transfer bound $\underline{\Delta} = [\underline{\Delta}_n^m]$ which gives, for each payoff function F^m, the lowest additive function above F^m and touching F^m at $\hat{x}$. We call such an additive function the *additive upper envelope* at $\hat{x}$ of the principals' excess payoffs F^m, $m \in \mathcal{M}$. We now give an explicit expression of $\underline{\Delta}$ for a class of principal excess payoffs.

PROPOSITION 8 (ADDITIVE UPPER ENVELOPE) *Suppose that principal m's excess payoff function $F^m : \mathcal{X} \to \mathbb{R}$ is of the form*

$$F^m(x_1, \ldots, x_N) = f^m\left(g_1^m(x_1), \ldots, g_N^m(x_N)\right),$$

where f^m is submodular[16] *and vanishes at $(0, \ldots, 0)$, and where the function $g_n^m : \mathcal{X}_n \to \mathbb{R}_+$ vanishes at $\hat{x}_n$ for all $n \in \mathcal{N}$. Then the unique additive upper envelope of F^m whose elements $\{\underline{\Delta}_n^m\}_{n\in\mathcal{N}}$ respectively vanish at $\{\hat{x}_n\}_{n\in\mathcal{N}}$ is*

$$\underline{\Delta}_n^m(x_n) = f^m(0, \ldots, 0, g_n^m(x_n), 0, \ldots, 0).$$

[16] A function h of k variables $(x_1, \ldots x_k) \in \mathbb{R}^{qk}$ where q is a positive integer is *submodular* if and only if for any $(x, y) \in \mathbb{R}^{2qk}$, $h(\min(x, y)) + h(\max(x, y) \leq h(x) + h(y)$, where the minimum and maximum are taken componentwise.

Proof. Submodularity of f^m, nonnegativity of the functions g_n^m, and the facts that f^m vanishes at $(0, \ldots, 0)$ and for each n, g_n^m vanishes at $\hat{x}_n$, imply that

$$F^m(x_1, \ldots, x_N) = f^m\left(g_1^m(x_1), \ldots, g_N^m(x_N)\right) \le \sum_{n \in \mathcal{N}} f^m(0, \ldots, 0, g_n^m(x_n), 0, \ldots, 0).$$

Moreover, letting $x_j = \hat{x}_j$ for $j \neq n$, and using the fact $g_i^m(\hat{x}_i) = 0$, yields

$$F^m(\hat{x}_1, \ldots, \hat{x}_{n-1}, x_n, \hat{x}_{x+1}, \ldots, \hat{x}_N) = f^m(0, \ldots, 0, g_n^m(x_n), 0, \ldots, 0),$$

which implies that

$$f^m(0, \ldots, \hat{0}, g_n^m(x_n), 0, \ldots, 0) \le \Delta_n^m(x_n) \tag{42}$$

for any admissible excess transfer matrix Δ, $n \in \mathcal{N}$, and $x_n \in \mathcal{X}_n$. Therefore, the excess transfers

$$\Delta_n^m(x_n) = f^m(0, \ldots, 0, g_n^m(x_n), 0, \ldots, 0) \tag{43}$$

for all $n \in \mathcal{N}$ and $x_n \in \mathcal{X}_n$ constitute the additive upper envelope of F^m. ■

A useful application of Proposition 8 is given in the following corollary.

COROLLARY 1 *Suppose that principal m's excess payoff function $F^m : \mathcal{X} \to \mathbb{R}$ is of the form*

$$F^m(x_1, \ldots, x_N) = f^m\left(c^m + \sum_{n \in \mathcal{N}} g_n^m(x_n)\right),$$

where $f^m : [c^m, \infty) \to \mathbb{R}$ is concave and vanishes at c^m, and where the function $g_n^m : \mathcal{X}_n \to \mathbb{R}_+$ vanishes at $\hat{x}_n$ for all $n \in \mathcal{N}$. Then, the unique additive upper envelope of F^m whose elements $\{\underline{\Delta}_n^m\}_{n \in \mathcal{N}}$ respectively vanish at $\{\hat{x}_n\}_{n \in \mathcal{N}}$ is

$$\underline{\Delta}_n^m(x_n) = f^m(c^m + g_n^m(x_n)).$$

Proof. Letting

$$\tilde{f}^m(y_1, \ldots, y_N) = f^m\left(c^m + \sum_{n \in \mathcal{N}} y_n\right), \tag{44}$$

concavity of f^m insures that $\tilde{f}^m$ is submodular.[17] Therefore, Proposition 8 can be applied to

$$F^m = \tilde{f}^m(g_1^m(x_1), \ldots, g_N^m(x_m)), \tag{45}$$

which yields $\underline{\Delta}_n^m = \tilde{f}^m(0, \ldots, 0, g_n^m(x_n), 0, \ldots, 0) = f^m(c^m + g_n^m(x_n))$. ■

EXAMPLES. Suppose that the set of outcomes is $\mathcal{X} = [0, \bar{x}] \times [0, \bar{y}]$ for some positive constants $\bar{x}, \bar{y}$, that the efficient outcome is $(\hat{x}, \hat{y}) = (0, 0)$, and that $F^m(x, y) = \ln(1 + k_1^m x + k_2^m y)$, with $k_i > 0$. Then, the additive upper envelope of the payoff function F^m vanishing at $(0,0)$ is $\underline{\Delta}_1^m(x) = \ln(1 + k_1^m x)$, and $\underline{\Delta}_2^m(y) = \ln(1 + k_2^m y)$. If we consider instead $F^m(x, y) = -(x^{k_1^m} + y^{k_2^m})^2$ with k_1^m, k_2^m positive, then the additive upper envelope is $\underline{\Delta}_1^m(x) = -x^{2k_1^m}$ and $\underline{\Delta}_2^m(y) = -y^{2k_2^m}$. Last, consider the constant-elasticity-of-substitution function $F^m(x, y) = [k_1^m x^\rho + k_2^m y^\rho]^{\frac{1}{\rho}}$, with k_1^m, k_2^m

[17]See Topkis (1968).

positive and $\rho > 1$. In this case, the additive upper envelope is $\underline{\Delta}_1^m(x) = (k_1^m)^{\frac{1}{\rho}} x$ and $\underline{\Delta}_2^m(y) = (k_2^m)^{\frac{1}{\rho}} y$. □

The examples illustrate the class of principal excess payoff functions for which a closed-form solution for the lower excess-transfer bound can be obtained. For this class of excess payoff functions, all WTEs of the game played through agents $\mathcal{G}$ can be obtained explicitly using the results in the preceding sections.

3.7 Discussion

We have shown that it is possible to construct weakly truthful equilibria for any game played through agents, as long as all principals' and agents' payoffs are concave.[18] The constructions can be used in a variety of practical settings with complete information, such as the coordination of multi-principal multi-agent supply chains. Note that no payoff externalities between agents are permitted in the model we discussed. The reason is that – as Segal (1999) shows – efficient outcomes may not be implementable when agents' payoffs depend on each others' actions. The situation also becomes more delicate when information about the contracts is asymmetric, as in bilateral contracting,[19] or when renegotiation is allowed, as in Matthews (1995). We have seen that any efficient outcome in games played through agents can be implemented with all principals using affine transfers, except for one principal. Nevertheless, to change the allocation of surplus within the principal-agent system for a given efficient outcome, it may be desirable to choose particular equilibrium excess

[18] We also provided conditions under which contracts implementing a weakly truthful equilibrium exist when payoffs are not concave.

[19] Segal and Whinston (2003) examine a setting where a single principal writes bilateral contracts with N different agents without announcing the contracts publicly.

transfer matrices. Using essentially convex combinations and a leveling algorithm, we have provided a class of admissible equilibrium excess transfers, which is complete under some assumptions. Our analysis of the relation between desired in-equilibrium surplus allocations α and off-equilibrium excess transfers Δ allows us to characterize and implement WTEs with Pareto-optimal in-equilibrium transfers. We have thus obtained a global solution to the following practical problem: given an efficient outcome, the principals determine the set of WTEs of $\mathcal{G}$. After agreeing on a point in the set of attainable Pareto-optimal in-equilibrium transfers (e.g., using a bargaining procedure) they implement the efficient outcome as a WTE. Bernheim and Whinston (1986) prove an equivalence between their truthful Nash equilibria and the concept of "coalition-proof Nash equilibrium" (see also Bernheim et al., 1997). This latter concept contains the idea that even if principals were allowed to collude, the corresponding Nash equilibria would still be outcomes of the games. Strengthening the concept of Nash equilibrium to coalition-proof equilibrium is therefore important for the predictive power of this theory whenever collusion among principals is possible or imperfectly monitored. While coalition-proofness is lost when using the more flexible concept of weakly truthful equilibrium, our characterization of Pareto optimal WTE may allow one to retrieve coalition-proof Nash equilibria. Indeed, coalition-proofness is equivalent to the property that in-equilibrium transfers are Pareto optimal, provided that the game is limited to any subset of principals, and that the remaining principals and all agents take their equilibrium actions.

Chapter 4

Performance-Sensitive Debt

This chapter studies *performance-sensitive debt* (PSD), the class of debt obligations whose interest payments depend on some measure of the borrower's performance. For example, step-up bonds compensate credit rating downgrades with higher interest rates, and reward credit rating upgrades with lower interest rates.[1] Performance pricing loans, a large fraction of commercial loans, also tie their interest rates to some measure of the borrower's credit quality.[2] In an endogenous default setting, we develop an algorithm to value PSD obligations allowing for general payment profiles, and obtain closed-form pricing formulas in important special cases, including step-up bonds. Moreover, we provide a criterion to compare different PSD obligations in terms of their efficiency. In particular, we find that step-up bonds lead to earlier default and lower market value of the issuing firm, compared to fixed-coupon bonds of the same market value. Lastly, we analyze the implications of our results for the policy of credit-rating agencies.

[1]Step-up bonds exceed \$100bn for both US- and European-based issuers (see Lando and Mortensen (2003) and "Step lightly," CFO Magazine (January 2001)).

[2]These loans represent over 70% of commercial loans (see Asquith, Beatty, and Weber (2002)).

4.1 Introduction

PSD obligations, including step-up bonds and performance pricing loans, compensate debtholders for changes in the borrower's credit risk. Practitioners have not yet reached any consensus on the likely effects of these *risk-compensating* PSD schemes. While proponents laud their high-yield, low-volatility characteristics (some even finding them "too generous"[3]), critics argue that risk-compensating PSD schemes generate a vicious circle by increasing the burden of debt service during financial strains, harming the issuer even more and, eventually, harming investors.[4] Underlying this disagreement is the lack of a theoretical model to value PSD and to assess the effect of issuing PSD rather than standard debt. This latter difficulty can be formalized as follows: for a given amount of debt raised, risk-compensating PSD results in paying higher interest than standard debt in times of low performance, and lower interest in times of high performance. It is unclear, then, between the perspective of lighter debt burden in times of high performance and the increased payment strains in times of low performance, which type of debt is more desirable.

Our goal is to build a valuation model for PSD, and to investigate how different types of PSD affect the timing of default and the equity value of the issuing firm. We develop a pricing algorithm allowing, tractably, for general payment profiles. We show that the equity value associated with PSD satisfies an ordinary differential equation with a boundary condition corresponding to zero value at default, and a "smooth-pasting" condition. We obtain closed-form pricing of PSD in important special cases, including step-up bonds. Building on our valuation model, we find that risk-compensating

[3] "The price of protection," Credit Magazine (September 1st, 2002)
[4] "Credit ratings can harm your wealth," Investment Adviser (December 9th, 2002).

PSD schemes have an overall negative effect on the issuing firm. In particular, issuing risk-compensating PSD leads to earlier default and, consequently, lowers the market value of the issuing firm's equity, holding constant the amount of cash raised by the obligation. Our results also bear implications on the behavior of credit-rating agencies. In trying to avoid the "credit-cliff dynamic", rating agencies are sometimes reluctant to downgrade distressed firms with PSD obligations in their capital structure.[5] Reluctant agencies generate distortions between actual and theoretical ratings, affecting the reliability of credit rating agencies.

Models of the valuation of risky debt can be divided into two classes. The first class treats firm's liabilities as contingent claims on its underlying assets, and bankruptcy as an endogenous decision of the firm. This set includes Black and Cox (1976), Fischer, Heinkel, and Zechner (1989), Leland (1994), Leland and Toft (1996) and Duffie and Lando (2001). In the second class of models, bankruptcy is not a decision of the firm. There is either an exogenous default boundary for the firm's assets (see Merton (1974) and Longstaff and Schwartz (1995)), or an exogenous process for the timing of bankruptcy, as in Jarrow and Turnbull (1995), Jarrow, Lando, and Turnbull (1997), and Duffie and Singleton (1999). Das and Tufano (1996), Acharya, Das, and Sundaram (2002), Houweling, Mentink, and Vorst (2003) and Lando and Mortensen (2003) obtain pricing formulas for credit-sensitive notes using the second class of models of the valuation of risky debt. Since they consider an exogenous default process, the costs associated with performance-sensitive debt do not become apparent in their models.

[5]See Standard and Poor's (2001).

In order to assess these costs, we work in the setting of Leland (1994), in which default is an endogenous decision of the firm. Instead of a fixed-coupon consol bond (paying a fixed interest rate), we consider debt obligations in which the interest rate is linked to some performance measure of the borrower. Performance-sensitive debt is thus fully characterized in this setting by some $C : \Pi \mapsto \mathbb{R}_+$ that maps a performance measure π lying in an ordered set Π to the interest rate $C(\pi)$ charged on the debt. Typical performance measures are credit ratings and financial ratios such as debt-to-earnings, leverage, or interest coverage. For PSD obligations C and D that are based on the same performance measure, we say that C is *more risk-compensating* than D if $C - D$ is nonincreasing and nonconstant. We prove that if C and D raise the same amount of cash, and if C is more risk-compensating than D, then C is less efficient than D, in the sense that it induces an earlier default time, therefore a higher present value of bankruptcy costs, and thus reduces the initial market value of the issuing equity. Therefore, it turns out that the trade-off between the opposite effects of the more risk-compensating scheme – relatively higher coupons in times of low performance and lower coupons in times of high performance – is systematically resolved in favor of the less risk-compensating debt.[6] We propose the following interpretation for this result. At the time of default, the more risk-compensating PSD requires higher interest payments, increasing the firm's losses. Although it is possible that this PSD imposes a lighter debt burden in the future, the current situation has a higher weight on equityholders' decision, and makes it less attractive for them to continue running the firm.

[6]This result is somewhat related to the finding by Hillion and Vermaelen (2004) that the issuance of floating-priced convertibles is followed by significant negative abnormal returns. Hillion and Vermaelen point out that the design of floating-priced convertibles encourages speculative short-selling activities by the convertible holders that can hurt the shareholders. This chapter does not consider convertibles or market speculation.

The remainder of the chapter is organized as follows. In Section 4.2, we illustrate several applications. In Section 4.3, we present the general model and formalize the notion of PSD. Section 4.4 analyzes the case of asset-based PSD obligations, demonstrating their relative efficiency. In Section 4.5, we explicitly derive the valuation of step-up and linear PSD obligations. Section 4.6 discusses different performance measures used in practice, and solves for the case of ratings-based PSD. Section 4.7 discusses the implications of our results for rating agencies policy. Section 4.8 provides additional discussion. Section 4.9 concludes.

4.2 Applications of PSD

This section describes PSD obligations that arise in practice. Some types of PSD obligations, such as credit-sensitive notes, performance-pricing loans and catastrophe bonds, have explicit performance-pricing provisions. Other types of PSD obligations are implicitly performance dependent because the terms of the debt are subject to renegotiation or are the result of an optimal dynamic capital strategy.

Credit-sensitive notes. A credit-sensitive note, sometimes called a step-up bond, pays an interest rate that is contractually linked with the credit rating of the borrower. First issued in the late 1980s, credit-sensitive notes have recently experienced an upsurge, specially among European telecommunications companies.[7]

Performance-pricing loans. Performance-pricing loans explicitly tie their interest to some pre-specified performance measure of the borrower. Typical performance

[7]Houweling, Mentink, and Vorst (2003) and Lando and Mortensen (2003) study the pricing of recent European telecommunications step-up bonds.

measures used for this purpose are credit ratings and such financial ratios as debt-to-earnings, leverage, or interest coverage. In an analysis of the Loan Pricing Corporation Database, Asquith, Beatty, and Weber (2002) found that the proportion of lending agreements including performance pricing provisions covered by this database increased from 40% in 1994 to over 70% in 1998.

Put-call provisions. Suppose a debt issue has provisions allowing the lending bank to put the debt back to the issuer when some performance measure drops below a contractual threshold. When such a provision is triggered, the lending bank often renegotiates the initial terms of the loan in effect, increasing the interest rate. The borrower may be given an option to call the loan when its credit quality improves. This permits the borrower to refinance the debt at lower interest rates after good performance. The outcome of these forms of optionality is effectively PSD.

Reset bonds. A reset bond, sometimes called a payment-in-kind (PIK) bond, has an interest rate that is adjusted periodically so that the market value of the bond is the same as its principal. In some cases the new interest rate is determined by an auction. The associate coupon rate C is thus decreasing in the credit quality of the borrower and a reset bond is, in effect, a form of PSD. Default in the junk-bond market may be induced by the rise in coupon payments of reset bonds.[8]

Short-term debt. The simplest case of PSD is short-term debt, such as commercial paper, since the coupon rate rises and falls continuously with the credit quality of the borrower. Myers (1977) argues that short-term debt may be used to mitigate agency costs. In Diamond (1991), risky firms do not issue short-term debt in order to avoid

[8] "The Junk-Bond Time Bombs Could Go Off," Business Week (April 9th, 1990).

early liquidation. Guedes and Opler (1996) provide empirical evidence supporting both claims.

Catastrophe bonds. Catastrophe (CAT) bonds, usually issued by insurance companies, promise coupons that are reduced in case total losses in the insurance industry exceed a pre-specified threshold.[9]

Dynamic capital structure. In a setting with taxes and bankruptcy costs, the optimal amount of debt outstanding varies with asset level. When the asset level increases, for example, issuers are better off by issuing more debt, since this gives them higher tax benefits. On the other hand, when the asset level decreases, debt reductions are optimal, ignoring transaction costs, as they reduce the present value of bankruptcy costs. The net effect, under some conditions, is PSD. This setting is studied in Goldstein, Ju, and Leland (1998).

4.3 The General Model

We begin by specifying a general model. Further assumptions will be added in later sections. We consider a generalization of the optimal liquidation models of Fischer, Heinkel, and Zechner (1989) and Leland (1994).[10] A firm generates cash flows at the rate δ_t, at each time t. We assume that δ is a diffusion defined by

$$d\delta_t = \mu_\delta(\delta_t)dt + \sigma_\delta(\delta_t)dB_t, \tag{1}$$

[9] See Fitch IBCA (2001) for a survey of the market for CAT bonds.

[10] While previous model specifications are limited to geometric Brownian motion, we consider here a general diffusion model.

where μ_δ and σ_δ satisfy the classic assumptions for the existence of a unique strong solution[11] to (1) on a fixed probability space $(\Omega, \mathcal{F}, P)$ with the information filtration $(\mathcal{F}_t)$ generated by the standard Brownian motion B. For simplicity, we assume that all agents are risk-neutral. There is a constant risk-free interest rate r, with $\mu_\delta < r$. The market value A_t at time t of the future cash flows of the firm is then

$$A_t = E_t \left[\int_t^\infty e^{-r(s-t)} \, \delta_s \, ds \right] < \infty \tag{2}$$

where E_t denotes the $\mathcal{F}_t$-conditional expectation. By the Markov property, A_t only depends on $\{\delta\}_{s \leq t}$ through δ_t. Specifically, there exists a smooth function $A : \mathbb{R} \to \mathbb{R}$ such that $A_t = A(\delta_t)$, which implies that $\{A_t\}_{t \geq 0}$ is a diffusion:

$$dA_t = \mu(A_t)dt + \sigma(A_t)dB_t. \tag{3}$$

For the sake of ulterior computations, we impose the following condition.

CONDITION 1 *μ and σ are smooth and bounded and σ is coercive.*[12]

Since $E_t[\delta_s]$ is increasing in δ_t, $A(\cdot)$ is increasing, which implies the existence of a smooth cash-flow rate function $\delta : \mathbb{R} \to \mathbb{R}$ such that $\delta_t = \delta(A_t)$.

We consider a performance measure represented by an $\mathcal{F}_t$-adapted stochastic process $(\pi_t)_{0 \leq t < \infty}$ taking values in a totally ordered, topological space Π. In general, π_t can be any statistic that measures the firm's ability and willingness to serve its debt obligations in the future. Financial ratios and credit ratings are among commonly used

[11] See for example Karatzas and Shreve (1991)

[12] In the one-dimensional case, coerciveness means that there exists a real number $\underline{\sigma}$ such that $0 < \underline{\sigma} \leq \sigma$.

performance measures. A *performance-sensitive debt* (PSD) obligation is a claim on the firm that promises a nonnegative payment rate that may vary with the performance measure of the firm. Formally, a PSD obligation $C(\cdot)$ is a (time-invariant) measurable function $C : \Pi \rightarrow \mathbb{R}$, such that the firm pays $C(\pi_t)$ to the debtholders at time t.[13] For example, the consol bond of Leland (1994) is a degenerate case of PSD. The reader should note that, while our earlier sections dealt mostly with "risk-compensating" PSD (that pay higher coupons when performance worsens), our definition of PSD encompasses more general kinds of PSD.

Given a PSD obligation C, the firm's optimal liquidation problem[14] is to choose a default time $\widehat{\tau}$ to maximize its initial equity value W_0^C, given the debt structure C. That is,

$$W_0^C \equiv \sup_{\widehat{\tau} \in \mathcal{T}} E \left[\int_0^{\widehat{\tau}} e^{-rt} [\delta_t - (1-\theta) C(\pi_t)] \, dt \right] , \tag{4}$$

where $\mathcal{T}$ is the set of $\mathcal{F}_t$ stopping times and θ is the corporate tax rate on earnings. If τ^* is the optimal liquidation time, then the value of equity at time $t < \tau^*$ is

$$W_t^C = E_t \left[\int_t^{\tau^*} e^{-r(s-t)} [\delta_s - (1-\theta) C(\pi_s)] \, ds \right] . \tag{5}$$

[13] We are considering perpetual debt, which is a standard simplifying assumption for the endogenous default framework. See, for example, Leland (1994). However, our model can be extended to the case of finite average debt maturity, if we assume that debt is continuously retired at par at a constant fractional rate. See Leland (1998) for more on this approach. Leland and Toft (1996) use more general finite-maturity debt framework. However, due to the complexity of their model, most of their results are obtained using simulations.

[14] Firms usually have standard fixed-coupon bonds together with different types of PSD obligations in their capital structure. In this case, the total outstanding debt of the firm is still PSD, but one has to sum the payment rates of all debt obligations issued by the firm when determining the total payment rate of the firm, which is the relevant payment rate for liquidation purposes.

Analogously, the market value U_t^C of the PSD obligation C at time t is

$$U_t^C \equiv E_t\left[\int_t^{\tau^*} e^{-r(s-t)}C(\pi_s)\,ds\right] + E_t\left[e^{-r(\tau^*-t)}\left(A_{\tau^*} - \rho(A_{\tau^*})\right)\right]\ , \tag{6}$$

where the function ρ defines the portion of the asset value lost at bankruptcy. The value of debt, as expressed by (6), has two components: the expected discounted coupon payments and the expected lump repayment at time default. We assume that ρ is an increasing function such that $0 \leq \rho(x) \leq x$ for all $x \geq 0$. If δ_t is lower than $(1-\theta)C(\pi_t)$, equity holders have a net negative dividend rate.[15] Equity holders will continue to operate a firm with negative dividend rate if the firm's prospects are good enough to compensate for the temporary losses.

4.4 Asset-Based PSD

In all the applications of PSD listed in the Section 4.2, the interest rate charged to the borrower depends on the borrower's credit quality. Since the market value A of assets is a time-homogeneous Markov process, the current asset level A_t is the only state variable in our model, and any measure of the borrower's earnings prospect at time t is solely determined by A_t. Therefore, it is natural to consider the asset level A_t as a performance measure. An *asset-based PSD* is a PSD whose coupon rate only depends on the current asset level. Specifically, an asset-based PSD is a measurable function $C : \mathbb{R}_+ \to \mathbb{R}$, under which the firm pays coupons at rate $C(A_t)$ at time t. Using this definition, we derive valuation and efficiency results for asset-based PSD.

[15]Limited liability is satisfied if the negative dividend rate is funded by dilution, for example through share purchase rights issued to current shareholders at the current valuation.

4.4.1 Valuation

Given an asset-based PSD, the initial value of the equity is:[16]

$$W(A_0) \equiv \sup_{\hat{\tau} \in \mathcal{T}} E\left[\int_0^{\hat{\tau}} e^{-rt}\left[\delta(A_t) - (1-\theta)C\left(A_t\right)\right] \, dt\right].$$

The Markov property and time homogeneity imply that there exist asset levels A_B and A_H with $A_B < A_0 < A_H$, such that an optimal default time of the firm is of the form $\tau^* = \min\left\{\tau(A_B), \tau(A_H)\right\}$, where $\tau(x) \equiv \inf\left\{t : A_t = x\right\}$. Even though the existence of an upper asset boundary A_H above which the firm would default is mathematically possible, we exclude this unnatural possibility with the following condition.

Condition 2 *There exist levels $\underline{x} < \bar{x}$ and a positive constant $\underline{c}$ such that*

1. $(1-\theta)C(x) \geq \delta(x)$ *if and only if* $x \leq \bar{x}$.
2. $(1-\theta)C(x) \geq \delta(x) + \underline{c}$ *for* $x \leq \underline{x}$.

The first part of Condition 2 states that for asset levels higher than $\bar{x}$, the cash flow rate is higher than the coupon rate. It can be easily verified that, under this condition, $A_H = +\infty$, so that the optimal default time simplifies to $\tau^* = \tau(A_B)$. Therefore, equity holders' optimal stopping problem can be expressed without loss of generality as:

$$W(x) = \sup_{y<x} \tilde{W}(x,y), \tag{7}$$

where

$$\tilde{W}(x,y) \equiv E_x\left[\int_0^{\tau(y)} e^{-rt}\left[\delta(A_t) - (1-\theta)C(A_t)\right] dt\right].$$

[16] Throughout this section, we omit the superscript C and the subscript 0 whenever there is no ambiguity.

In order to derive an ordinary differential equation (ODE) for W, we impose the following condition on C:

CONDITION 3 *The PSD obligation C is such that:*

1. *There exist nonnegative constants k_1 and k_2 that satisfy*

$$0 \leq (1-\theta)C(y) \leq k_1 + k_2 y.$$

2. *C is right-continuous on $[0,\infty)$ and has left limits on $(0,\infty)$.*

Using the strong Markov property of $\{A_t\}_{t\geq 0}$,

$$\tilde{W}(x,y) = f(x) - \xi(x,y)f(y) \tag{8}$$

where[17]

$$\xi(x,y) = E_x[e^{-\tau(y)}],$$

and

$$f(x) = E_x\left[\int_0^\infty e^{-rt}\left[\delta(A_t) - (1-\theta)C(A_t)\right] dt\right].$$

The next lemma shows that, under Condition 2, the default triggering level A_B is strictly positive.

LEMMA 1 *Under Condition 2, there exists a level $\tilde{x}$ such that any optimal default time τ satisfies $\tau \leq \tau(\tilde{x})$ almost surely.*

An important consequence of Lemma 1 is that default occurs with positive probability. Our next theorem characterizes the solution of the optimal stopping problem (7).

[17]Previous assumptions on μ and σ imply that ξ is well-defined (see Karatzas and Shreve (1991))

THEOREM 1 *If a PSD C satisfies Conditions 1–3, the following statements are equivalent:*

1. *A_B is an optimal default triggering level:*

$$W(x) = E_x\left[\int_0^{\tau(A_B)} e^{-r(s-t)}[\delta(A_s) - (1-\theta)C(A_s)]\,ds\right].$$

2. *$W(x)$ and A_B satisfy:*

   (i) $A_B \in (0, \bar{x})$.

   (ii) *W is continuously differentiable and W' is bounded and left and right differentiable.*

   (iii) *W vanishes on $[0, A_B]$ and satisfies the following ODE at any point of continuity of C:*

$$\frac{1}{2}\sigma^2(x)W''(x) + \mu(x)W'(x) - rW(x) + \delta(x) - (1-\theta)C(x) = 0. \quad (9)$$

A proof is given[18] in Section 4.10.

The continuous differentiability of W and the fact that W is 0 on $[0, A_B]$ imply that $W'(A_B) = 0$, which is known as the *smooth-pasting condition.* Theorem 1 provides a method for solving the firm's optimal liquidation problem. The proposed algorithm is the following

[18] The Appendix also gives two separate equations involving the right and left derivatives of W' at discontinuity points of C (cf. equations (25) and (26)).

1. Determine the set of continuously differentiable functions that solve ODE (9) at every continuity point of C. It can be shown that any element of this set can be represented with two parameters,[19] say L_1 and L_2.

2. Determine A_B, L_1, and L_2 using the following conditions:

a. $W(A_B) = 0$.

b. W' is bounded.

c. $W'(A_B) = 0$.

d. $A_B \in (0, \bar{x})$.

We interpret (a) as the boundary condition on the solution at the point A_B of the ODE. Condition (b) says that $W'(x)$ remains bounded as $x \to +\infty$ and constitutes the second boundary condition on the solution of the ODE. The smooth-pasting condition (c) is interpreted as the first-order optimization condition that defines the optimal bankruptcy boundary. Condition (d) verifies that condition 2.(i) of Theorem 1 is satisfied.

Now that we know how to price the equity associated with PSD, we can also price the PSD itself. Using the fact that the sum of the equity value, the PSD value, and the expected losses resulting from the bankruptcy is the sum of the asset level and the present value of the tax benefits, we obtain the PSD pricing formula:

$$U(A_t) = \frac{1}{1-\theta}\left[A_t - W(A_t) - [\rho(A_B) + \theta(A_B - \rho(A_B))]\,\xi(A_t, A_B)\right].$$

[19] In fact, we really consider here solutions of coupled equations (25) and (26), which boil down to the ODE (9) at any continuity point of C. One can easily check that the set of solutions of the coupled equations still is a two-dimensional vector space.

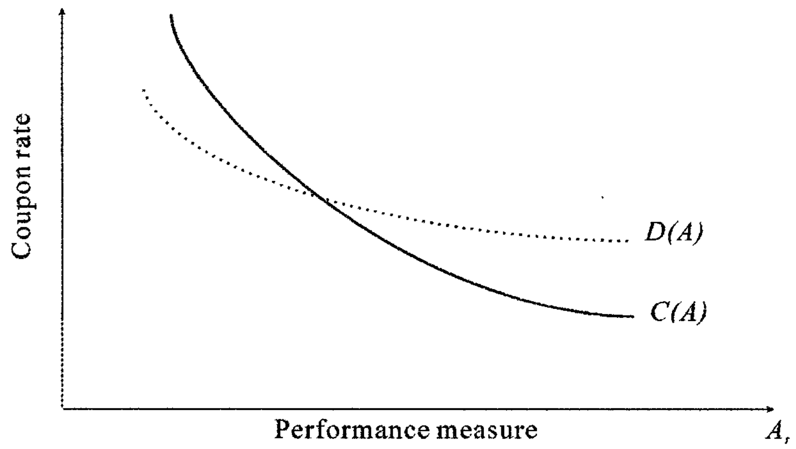

Figure 1: C is more risk compensating than D.

4.4.2 The Relative Efficiency of Asset-Based PSD

In this subsection, we study the relative efficiency of alternative asset-based PSD. Specifically, we derive a partial order, by "efficiency," among alternative PSD issues that raise the same amount of cash. We need the following definitions and condition, that we state in terms of a general performance measure π. These will also be used in Section 4.6.2, for the case of credit ratings.

DEFINITION 1 (Relative Efficiency). *Let C and D be PSD that raise the same funds, $U_0^C = U_0^D$. We say that C is* less efficient *than D if it determines a lower equity price, that is, if $W_0^C < W_0^D$.*

DEFINITION 2 (Risk Compensating). *Let C and D be PSD obligations based on the same performance measure. We say that C is* more risk-compensating *than D if $C - D$ is a nonincreasing, nonconstant function.*

Figure 1 illustrates the "risk compensating" concept.

CONDITION 4 (Efficiency Domain). *A PSD issue C is said to be in its efficiency domain if, for any constant $\alpha > 0$, we have $U_0^{C-\alpha} < U_0^C$, where $C - \alpha$ denotes a PSD issue that pays $C(A_t) - \alpha$ at time t.*

Condition 4 means that it is not possible to raise the same amount of cash as C by a constant downward shift in its coupon rate. For example, a bond paying a fixed coupon rate c raises an increasing amount of cash as c increases, until c reaches a point at which the loss due to precipitated default dominates the gain due to the increase of coupon payment (as in Figure 2). The forms of PSD that we consider are in their efficiency domain, for otherwise efficiency in the sense of Definition 1 can be trivially improved upon by uniformly reducing the interest rate paid.

THEOREM 2 *Suppose C and D both are asset-based PSD, satisfying $U_0^C = U_0^D$ and Conditions 1–4. If C is more risk-compensating than D, then C is less efficient than D.*

A proof of the theorem is given in Section 4.10.

The above result is supported by the following intuition. Equityholders decide to declare bankruptcy when coupon payments become too high compared to dividends. At this time, the firm pays higher interest rates with C than with D and, while there is a possibility that the situation be reversed in the future, the urgency of the current situation increases the firm's incentive to declare bankruptcy. This intuition can be further illustrated by the opposite, extreme example of a bond paying a coupon rate equal to the dividend rate $C(A_t) = \delta(A_t)$. This coupon rate decreases to zero as the asset level goes to zero. The coupon payments never exceed the dividends, so the firm never goes bankrupt. Such a bond transfers all the value of the firm to debtholders,

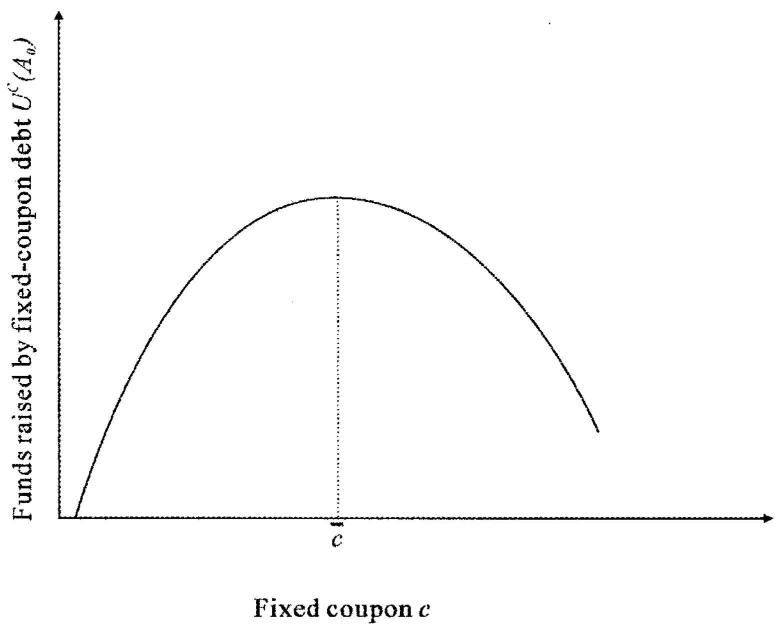

Figure 2: A fixed-coupon bond is in its efficiency domain if $c \in [0, \bar{c}]$.

and, if it could qualify as "debt" for tax purposes, would reduce tax payments to zero since the tax benefit resulting from coupon payments is equal to the tax on the dividends. Equityholders could decide to buy all of the debt, in which case this bond allows them to receive all of their dividends in form of coupon payments.

COROLLARY 1 *Let C be a PSD issue satisfying Conditions 1–4. If C is non-increasing and not constant, it is less efficient than the fixed-interest PSD issue raising the same amount of cash and verifying Condition 4. If C is non-decreasing and not constant, it is more efficient than any fixed-interest PSD issue raising the same amount of cash.*

The result suggests that, in many settings, the issuer would choose the least risk-compensating form of debt that qualifies as "debt" for tax treatment.

0

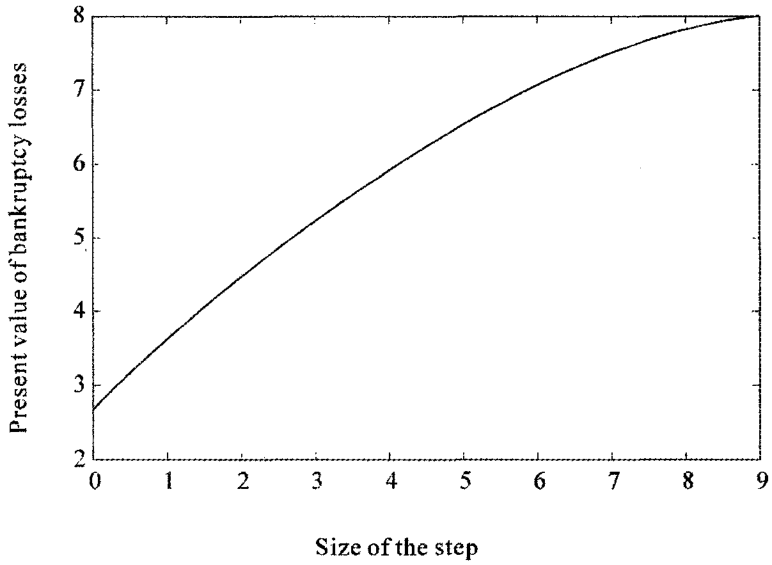

Figure 3: Present value of bankruptcy costs as a function of the step size.

The following numerical example compares "one-step" PSD issues C that raise the same amount M, in the class $\mathcal{C}_M$ of PSD defined by

$$C(A_t) = \begin{cases} C_1, & A_t \geq G_2, \\ C_2, & A_t < G_2, \end{cases}$$

such that $C_2 \geq C_1$ and $U^C(A_0) = M$.

We assume that the asset is a geometric Brownian motion with parameters $\mu = 0.01$, $\sigma = 0.1$, and that $\rho(x) = 0.25x$, $\theta = 0$, $r = 0.03$, $A_0 = 100$, $G_2 = 80$, and $M = 50$. M can be raised by issuing a bond that promises to pay a fixed coupon rate of 2. To see the inefficiency of step-up bonds, we compute for one-step PSD issues in $\mathcal{C}_M$ the present value of bankruptcy losses, which is by definition

$$Q(C) \equiv 0.25E\left[e^{-r\tau(A_B^C)}A_B^C\right] = 0.25A_B^C\left(\frac{A_0}{A_B^C}\right)^{-\gamma_1},$$

where[20] $\gamma_1 = \dfrac{m + \sqrt{m^2 + 2r\sigma^2}}{\sigma^2}$ and $m = \mu - \frac{\sigma^2}{2}$. According to Definition 2, (C_1, C_2) is more risk-compensating than (C_1', C_2') if $C_2 - C_1 > C_2' - C_1'$. Figure 3 shows the relationship between the present value of bankruptcy losses and the degree of risk-compensation $(C_2 - C_1)$ associated with the PSD. Expected bankruptcy costs equal 2.8 for the fixed-coupon PSD, and increase rapidly as step size takes off4.

4.5 Examples of Asset-Based PSD

In this section, we solve our model explicitly for two important cases: step-up and linear PSD issues. Step-up PSD is more likely to be seen in practice, while linear PSD has a convenient pricing formula. Throughout this section, we assume that the asset process is a geometric Brownian motion with drift μ and volatility σ^2. This implies that $\delta(x) = (r - \mu)x$, and that $\xi(x, y) = (x/y)^{-\gamma_1}$, where $\gamma_1 = \dfrac{m + \sqrt{m^2 + 2r\sigma^2}}{\sigma^2}$ and $m = \mu - \frac{\sigma^2}{2}$.

4.5.1 Step-Up PSD

Step-up performance-sensitive debt is defined to be a PSD obligation whose coupon payment is a non-increasing step function of the asset level.

For a decreasing sequence $\{G_i\}_{i=1}^{I+1}$ of asset levels such that $G_1 = +\infty$ and $G_{I+1} = A_B$,

[20]See Section 4.5

the coupon rate of a step-up PSD obligation can be represented as

$$C(A_t) = \bar{C}_i \text{ whenever } A_t \in [G_{i+1}, G_i) \; , \tag{10}$$

where $\{\bar{C}_i\}_{i=1}^{I}$ is an increasing sequence of constant coupon rates.

With this coupon structure, the general solution of the ODE (9) is

$$W(x) = \begin{cases} 0, & x \leq A_B \; , \\ L_i^{(1)} x^{-\gamma_1} + L_i^{(2)} x^{-\gamma_2} + x - \frac{(1-\theta)\bar{C}_i}{r}, & G_{i+1} \leq x \leq G_i \; , \end{cases} \tag{11}$$

for $i = 2, \ldots, I+1$, where $\gamma_1 = \dfrac{m + \sqrt{m^2 + 2r\sigma^2}}{\sigma^2}$, $\gamma_2 = \dfrac{m - \sqrt{m^2 + 2r\sigma^2}}{\sigma^2}$, $m = \mu - \dfrac{\sigma^2}{2}$, and where $L_i^{(1)}$ and $L_i^{(2)}$ are constants to be determined shortly. According to Theorem 1,

$$W(A_B) = 0 \tag{12}$$

and

$$W'(A_B) = 0 \; , \tag{13}$$

and $W(\cdot)$ is continuously differentiable. In particular, for $i = 2, \ldots, I$,

$$W(G_i-) = W(G_i+) \, , \; W'(G_i-) = W'(G_i+) \; . \tag{14}$$

Because the market value of equity is non-negative and cannot exceed the asset value[21],

$$L_1^{(2)} = 0 \; . \tag{15}$$

[21]Since $\gamma_1 > 0$ and $\gamma_2 < 0$, the term $L_K^2 x^{-\gamma_2}$ would necessarily dominate the other terms in the equation (11) violating the inequality $0 \leq W(x) \leq x$, unless $L_1^2 = 0$

The system (12)-(15) consists of $2I+1$ equations with $2I+1$ unknowns ($L_i^{(j)}$ with $j \in \{1,2\}$ and $i \in \{1,\ldots,I\}$, and A_B). Substituting (11) into (12)-(15) and solving gives

$$
\begin{aligned}
L_I^{(1)} &= \frac{(\gamma_2+1)\,A_B - \gamma_2 \frac{c_2}{r}}{(\gamma_1-\gamma_2)\,A_B^{-\gamma_1}}\,, && (16)\\
L_I^{(2)} &= \frac{-(\gamma_1+1)\,A_B + \gamma_1 \frac{c_2}{r}}{(\gamma_1-\gamma_2)\,A_B^{-\gamma_2}}\,, && (17)\\
L_i^{(1)} &= L_I^{(1)} + \frac{\gamma_2}{(\gamma_1-\gamma_2)r}\sum_{i=j}^{I-1}\frac{c_{i+1}-c_i}{G_{i+1}^{-\gamma_1}},\ i=2,\ldots,I\,, && (18)\\
L_i^{(2)} &= L_I^{(2)} - \frac{\gamma_1}{(\gamma_1-\gamma_2)r}\sum_{i=j}^{I-1}\frac{c_{i+1}-c_i}{G_{i+1}^{-\gamma_2}},\ i=2,\ldots,I\,, && (19)\\
0 &= -(\gamma_1+1)\,A_B + \frac{\gamma_1}{r}\left(c_I - \sum_{i=1}^{I-1}(c_{i+1}-c_i)\left(\frac{A_B}{G_{i+1}}\right)^{-\gamma_2}\right)\,, && (20)
\end{aligned}
$$

where, for convenience, we let $c_i \equiv (1-\theta)\bar{C}_i$. Although we do not have an explicit solution for these parameters, equations (16)-(19) express $L_j^{(i)}$ as a function of A_B, which, in turn, solves (20). One can verify that (20) has a unique solution on the interval $\left(0, \widehat{A}_B\right)$,[22] where $\widehat{A}_B \equiv \gamma_1 c_I/(r(\gamma_1+1))$ is the default-triggering level of assets for a consol bond with the fixed-coupon rate c_I.

[22]Since $\gamma_1 \in (0,\infty)$ and $\gamma_2 \in (-\infty,0)$, the left-hand-side of (20) converges to $\frac{\gamma_1}{r}c_I > 0$ as A_B goes to 0, and equals $-\frac{\gamma_1}{r}\sum_{i=1}^{I-1}(c_{i+1}-c_i)\left(\frac{A_B}{G_{i+1}}\right)^{-\gamma_2} < 0$ for $A_B = \widehat{A}_B$, where

$$\widehat{A}_B \equiv \frac{\gamma_1 c_I}{r(\gamma_1+1)}\,.$$

One can verify that the left-hand-side is a strictly decreasing function of A_B. Here, $\widehat{A}_B$ is the default-triggering level of assets for a consol bond with fixed-coupon c_I. Our step-up PSD pays several different coupon rates, and all of them are greater or equal than c_I. Therefore, A_B should be no greater than $\widehat{A}_B$, and (20) has a unique solution for A_B on the interval $\left(0, \widehat{A}_B\right)$.

4.5.2 Linear PSD

Consider the coupon scheme given by

$$C(x) = \beta_0 - \beta_1 x \ ,$$

with $\beta_0 > 0$. Applying Theorem 1, the corresponding equity value is

$$W(x) = \lambda\left(x - A_B\left(\frac{x}{A_b}\right)^{-\gamma_1}\right) - \frac{\beta_0}{r}\left(1 - \left(\frac{x}{A_B}\right)^{-\gamma_1}\right) \ , \tag{21}$$

and the optimal bankruptcy boundary is

$$A_B = \frac{\gamma_1 \beta_0}{\lambda(1+\gamma_1) r} \ ,$$

where $\lambda = \frac{r-\mu+\beta_1}{r-\mu}$. When $\beta_1 = 0$, $\lambda = 1$, formula (21) corresponds to the fixed coupon case with $C = \beta_0$. As expected, W is increasing in β_1 due to the reduction in the coupon rate.

4.6 Performance Measures

Earlier, we derived valuation formulas and an inefficiency theorem for PSD obligations whose coupon payments are determined by the asset level of the firm. Since, in our model, A_t incorporates all information about future earnings of the firm, the asset level is the natural choice for a performance measure. In practice, however, PSD contracts are usually written in terms of performance measures such as credit ratings and financial ratios. In this section, we explicitly consider the valuation and relative efficiency of PSD obligations based on these other performance measures.

4.6.1 General performance measures

We assume that performance measures reflect the borrower's capacity and willingness to repay the debt. Throughout this section, we assume that $\{A_t\}_{t\geq 0}$ follows a geometric Brownian motion with drift μ and volatility σ^2 (see Section 4.5 for the implications of this assumption). With μ and σ given, the borrower's asset level A_t and chosen default triggering boundary A_B fully determine its default characteristics at any time t. Since A_B is not directly observed by outsiders, the performance measure π_t is a function $\overline{\pi}(A_t, \tilde{A}_B)$, where $\tilde{A}_B$ is the perceived default triggering level of assets. Although we do not impose this condition, it is natural to think of $\overline{\pi}(\cdot,\cdot)$ as being nondecreasing in A_t and nonincreasing in $\tilde{A}_B$. A PSD obligation C therefore pays the coupon $C(\pi_t) = C(\overline{\pi}(A_t, \widetilde{A}_B))$. The Markov structure and the time homogeneity of the setting imply that any optimal default time of the firm can be simplified to a default triggering boundary hitting time $\tau(A_B)$ (still imposing Condition 2). In this setting, a consistency problem arises, as the default triggering level chosen by the firm may depend on the perceived default triggering level. With y denoting the actual default triggering level of the firm, the value of the equity is

$$\widetilde{W}(x, y, \widetilde{A}_B) = E_x\left[\int_0^{\tau(y)} e^{-rt}\left[(r-\mu)\,A_t - (1-\theta)C(\overline{\pi}(A_t, \widetilde{A}_B))\right] dt\right].$$

Knowing that the firm seeks to maximize the value of the equity, the ratings agency therefore chooses an $\widetilde{A}_B$ that solves the fixed point equation:

$$A_B \in \arg\max_{y\leq x} \widetilde{W}(x, y, A_B). \tag{22}$$

This equation may have one or several solutions, or no solution at all. To avoid ambiguity, we impose the following condition.

CONDITION 5 *There exists a unique positive solution of (22).*

Given Condition 5, the coupon rate paid by the PSD obligation at time t is $C(\overline{\pi}(A_t, A_B))$. Since A_B does not change over time, this PSD, which is defined under performance measure π, is equivalent to an asset-based PSD $\tilde{C}$, defined by $\tilde{C}(A_t) \equiv C(\overline{\pi}(A_t, A_B))$. Equation (22) implies that C and $\tilde{C}$ have the same optimal default boundary A_B. Hence, provided that $\tilde{C}$ satisfies Conditions 2,3, and 4, we can compare C in terms of efficiency with asset-based PSD obligations that satisfy the same Conditions by applying Theorem 2. In particular, if $\tilde{C}(A_t)$ is a nonincreasing nonnegative function, then a fixed-coupon bond with the same market value is more efficient than C. If π can only take finitely many values, then $\tilde{C}(A_t)$ satisfies Conditions 2 and 3. Thus, we have proved the following theorem.

THEOREM 3 *Suppose that a performance measure π can only take a finite number of values, and that a PSD C is nonincreasing and nonnegative. Suppose Conditions 4 and 5 are satisfied. Then, a fixed-coupon PSD D that satisfies Condition 4, and has the same market value as C ($U_0^C = U_0^D$), is more efficient than C.*

4.6.2 Ratings-based PSD

We consider I different credit ratings, $1, \ldots, I$, with 1 the highest ("Aaa" in Moody's ranking) and I the lowest ("C" in Moody's ranking). We let R_t denote the issuer's credit rating at time t. We say that $C \in \mathbb{R}^I$ is a ratings-based PSD obligation if it pays interest at the rate C_i whenever $R_t = i$, with $C_{i+1} \geq C_i > 0$, for i in $\{1, \ldots, I-1\}$. Thus, a *ratings-based PSD is more risk-compensating than a fixed coupon PSD.* We say that an *accurate* rating agency is one whose credit ratings are a function of the probability of default over a given time horizon T. Naturally, higher ratings correspond to lower default probabilities.

The default time for a ratings-based PSD is a stopping time of the form $\tau(A_B) = \inf\{s : A_s \leq A_B\}$, for some A_B. Therefore, the current asset level A_t is a sufficient statistic for $P(\tau(A_B) \leq T \mid \mathcal{F}_t)$, for any $T \geq t$. A rating policy is thus given by some vector-valued function $G : \mathbb{R} \mapsto \mathbb{R}^{I+1}$ that maps a default boundary A_B into rating transition thresholds, such that $R_t = i$ whenever $A_t \in [G_{i+1}(A_B), G_i(A_B))$. In our setting, this policy has the form[23]

$$G(A_B) = A_B\, g, \tag{23}$$

where $g \in \mathbb{R}^I$ is such that $g_1 = +\infty$, $g_{I+1} = 1$, and $g_i \geq g_{i+1}$.

The results developed for step-up PSD can be applied to ratings-based PSD. In particular, the maximum-equity-valuation problem (4) is solved by $\tau(A_B) = \inf\{s : A_s \leq A_B\}$, where A_B solves equation (20). Plugging (23) into (20), we obtain

$$A_B = \frac{\gamma_1}{(\gamma_1 + 1)r} \widehat{C}, \tag{24}$$

where

$$\widehat{C} = \sum_{i=1}^{I} \left[\left(\frac{1}{g_{i+1}} \right)^{-\gamma_2} - \left(\frac{1}{g_i} \right)^{-\gamma_2} \right] c_i,$$

[23]Since A_t is a geometric Brownian motion, its first passage time distribution is an inverse Gaussian:

$$P(\tau(A_B) \leq T \mid \mathcal{F}_t) = 1 - \Phi\left(\frac{m(T-t) - x}{\sigma\sqrt{T-t}} \right) + e^{\frac{2mx}{\sigma^2}} \Phi\left(\frac{x + m(T-t)}{\sigma\sqrt{T-t}} \right),$$

where, $x = ln\left(\frac{A_B}{A_t}\right)$, $m = \mu - \frac{1}{2}\sigma^2$, A_t is the current level of assets and Φ is the normal cumulative distribution function. Since $P(\tau(A_B) \leq T \mid \mathcal{F}_t)$ depends on A_t only through $\frac{A_B}{A_t}$, we have the linearity of $G(\cdot)$.

and $c_i = (1-\theta)C_i$. We note that the ratings-based PSD issue C has the same default boundary A_B as that of a fixed-coupon bond paying coupons at the rate $\widehat{C}$. Plugging (24) into (16)-(19), (11), and (6), one obtains closed-form expressions for the market value W of equity and the market value U of debt for any ratings-based PSD obligation.

We now derive the inefficiency theorem for the case of ratings-based PSD. We keep the same definitions as in Section 4.4, except that the performance measure now corresponds to credit ratings, and not asset levels.

THEOREM 4 *Suppose C and D are ratings-based PSD, satisfying $U_0^C = U_0^D$ and Condition 4. If C is more risk-compensating than D, then C is less efficient than D.*

The proof of the theorem is given in Section 4.10.

COROLLARY 2 *Let C be a ratings-based PSD issue satisfying Conditions 2,3, and 4. If C is not constant, it is less efficient than any fixed-interest PSD issue raising the same amount of cash and verifying Condition 3.*

4.7 Rating Agency Policy

Credit ratings differ from other measures because of the circularity issues that are imposed. In a ratings-based PSD obligation, the rating determines the coupon rate, which affects the optimal default decision of the issuer. This, in turn, influences the rating. We have so far assumed that rating agencies are accurate, in the sense that they assign credit ratings according to the probability of default over a time horizon T. In this section, we discuss what can happen when credit-rating agencies fail to account for the effect of credit-rating changes on the firm's financial standing. Only after recent deteriorations in credit qualities of several major companies did rating agencies begin to worry about the unintended adverse effects of rating triggers.[24] Even after several incidences of default and cascading downgrades related to ratings-based PSD, it is not difficult to find examples of reluctance by rating agencies to incorporate the negative consequences of ratings-based PSD into credit ratings.[25] The following passage is from Standard and Poor's (2001):

> (...) How is the vulnerability of rating triggers reflected all along in a company's ratings? Ironically, it typically is not a rating determinant, given the circularity issues that would be posed. To lower a rating because we might lower it makes little sense – specially if that action would trip the trigger!

Another reason that rating triggers may not be incorporated into credit ratings is that often, due to confidentiality constraints, they are not publicly disclosed by the issuer. Some steps have already been taken to punish issuers who refuse to provide information about their rating triggers, although there is still no legal procedure to enforce disclosure.[26] We say that an agency is *unresponsive* if it ignores, when assigning credit ratings, the adverse effects of rating triggers on the liquidation of the firm.

We suppose, for example, that a firm having a fixed-coupon note C refinances its outstanding debt by issuing a ratings-based PSD obligation D. Figure 4 plots the accurate agency policy $G(\cdot)$, which is obtained from (23), and equityholders' optimal

[24]See Moody's (2001) and Standard and Poor's (2001).
[25]Moody's adopted a more critical view of ratings trigger after recent default events. See Moody's (2001).
[26]See Moody's (2002).

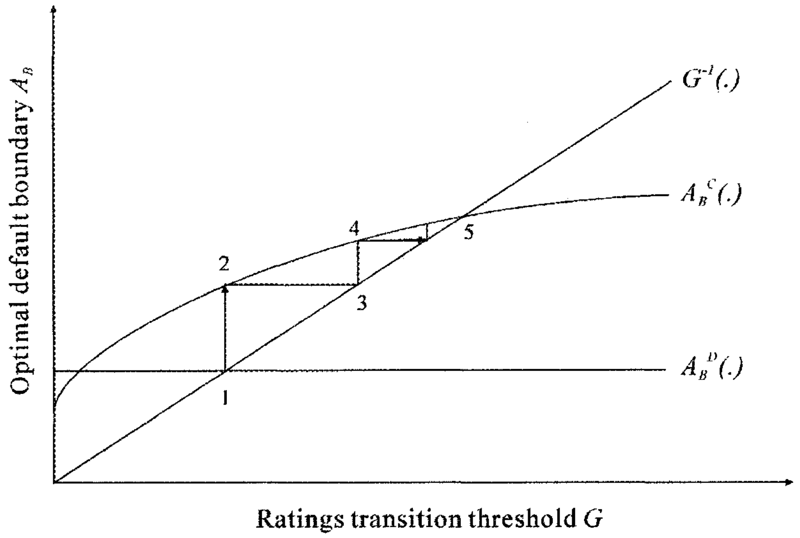

Figure 4: Present value of bankruptcy costs as a function of the step size.

default strategies $A_B^C(\cdot)$ and $A_B^D(\cdot)$, which are obtained from (20). Points 1 and 5 in the figure yield the solution to (22) before and after the refinancing of the debt takes place. With an accurate rating agency, issuance of ratings-based PSD obligations thus triggers a chain reaction that ceases only when it reaches point 5. This chain reaction, which we call *credit-cliff* dynamic, might induce a drastic downgrade or even immediate default if $A_B^C > A_0$. By ignoring the effects of ratings triggers, an unresponsive rating agency may avoid the perverse effects associated with the credit-cliff dynamic. In the context of figure 4, an unresponsive rating agency would interrupt the chain reaction at point 2, leading to a lower optimal default boundary than in the case of an accurate rating agency. One would then be tempted to say that the outcome of a ratings-based PSD with an unresponsive rating agency is superior to the one with an accurate one. We claim that this is not necessarily true. With unresponsive rating agencies, credit ratings do not reflect true probabilities of default and are thus less informative. Moreover, firms may be tempted to issue more risk-compensating ratings-based PSD, compensating for the unresponsiveness of rating agencies.

4.8 Additional Discussion

Even though our main result is that more risk-compensating PSD obligations lead to higher inefficiency, companies do issue these obligations in practice. In order to understand why this is the case, one could introduce market frictions such as adverse selection, moral hazard, contracting costs, or incomplete markets. Since these would complicate the model, we confine ourselves to an intuitive discussion of these issues.

Performance-sensitive debt may be used to solve the adverse selection problem, which arises because of information asymmetries at the time of debt issuance. In order to see this, we assume there are two firms that are identical except for their initial asset levels. That is, both firms' future cash flows are given by (1), but the "high" type has a higher initial level of assets than the "low" type. Assuming that their initial levels of assets are not observable by the market, the firm with the high assets may issue risk-compensating PSD that pays a lower initial coupon but has a higher associated bankruptcy boundary than that of the low-type firm that issued the fixed coupon debt. A lower asset level means that the firm is closer to bankruptcy. A further increase in the bankruptcy boundary would be costlier for the low-type firm. As a result, "low" type would not be willing to pool with the "high" type. On the other hand, despite the inefficiencies related to the risk-compensating PSD, the "high" type firm benefits overall from revealing its type by reducing its interest payments. Thus, the inefficiency cost associated with the risk-compensating PSD could be viewed as a signaling cost paid by the "high" type firm.[27]

Moral hazard could also justify the use of performance-sensitive debt. By punishing shareholders with higher interest rates after a bad performance, PSD obligations may reduce a manager's ability to shift wealth in favor of shareholders. We have solved a simple numerical example in which a firm that has access to high-risk and low-risk technologies issues step-up bonds in order to avoid losses from the asset-substitution effect.

Contracting costs may be another reason for some types of PSD. When the credit quality of the borrower changes, the issuer and the investors in its debt often get involved in costly negotiation over the terms of the debt. An increase in credit quality may prompt the borrower to seek refinancing of its debt on better terms. On the contrary, the lender may demand higher interest payments in compensation for the deterioration in credit quality. Some types of PSD may resolve the renegotiation problem by automatically adjusting the interest rates.

Asquith, Beatty, and Weber (2002) and Beatty, Dichev, and Weber (2002) indeed found empirical evidence that private debt contracts are more likely to include performance pricing schemes when asymmetric information, moral hazard or recontracting costs are significant. The chapter, however, establishes that solving these problems with PSD comes with a cost.

[27]My coauthors found numerical examples supporting this intuition.

We have so far assumed that all the agents in the economy are risk-neutral. It is straightforward, however, to extend our results to the case of risk-averse agents, in the absence of arbitrage (specifically, assuming the existence of an equivalent martingale measure).

If markets are incomplete, performance-sensitive debt might be issued to meet the demands of risk-averse investors, providing them with hedge against credit deterioration of the firm. Our results suggest, however that financial guarantors, rather than the debt issuing firms, should be providing this kind of hedge.

Our inefficiency results hold for different definitions of financial distress. If we assume, for example, that default happens when assets do not generate enough cash flow to meet current obligations[28], then it is easy to see that a more risk-compensating PSD will lead to more inefficiency. In this *flow-based* insolvency definition, however, shareholders declare bankruptcy even though it may be still possible to issue additional equity to cover the shortage.

[28]This setting is studied in Kim, Ramaswamy, and Sundaresan (1993).

4.9 Conclusion

This chapter analyzes the properties of performance sensitive debt using an endogenous default model. Although many types of debt contracts are performance-sensitive, they have received little attention in the literature. Endogeneity of the firm's default decision allows us to analyze the efficiency of different types of PSD. Our main finding is that, given the same initial funds raised by sale of debt, more risk-compensating PSD leads to earlier default and, consequently, lowers the market value of the issuing firm's equity. An intuitive explanation of this result is that higher interest payments from financially distressed companies lead to higher losses, thus precipitating the default decision. Catastrophe bonds, whose coupon rate is reduced automatically when the issuing insurance company experiences hardship due to a high volume of insurance claims, are an example of "more" efficient debt. The majority of PSD issues, however, have an inefficient step-up feature. This leads us to believe that inefficient PSD is used to solve agency problems arising from existing market imperfections, such as adverse selection, moral hazard and contracting costs. In addition, we develop a convenient method of valuing PSD. We obtain closed-form expressions for the equity prices associated with step-up, linear and rating-based PSD. We also discuss the policy of credit-rating agencies. Inconsistent rating of PSD can generate a credit-cliff dynamic, as well as hurt market participants by providing misleading information about default risks.

4.10 Appendix

Proof of Lemma 1. The proof is based on the following claim:

Claim: There exists a level $\tilde{x}$ such that $\forall x \leq \tilde{x}]$, $W(x) = \sup_{\tau} W(x, \tau) = 0$.

Proof. From Condition 2, there exist positive constants $\underline{x}$ and $\underline{c}$ such that $(1 - \theta)C(x) > \delta(x) + \underline{c}$ for all $x \leq \underline{x}$. Let $\Xi = \sup_{\tau} W(\underline{x}, \tau) < \infty$. For any stopping time τ and $x < \underline{x}$,

$$\begin{aligned} W(x,\tau) &= E_x\left[1_{\tau<\tau(\underline{x})} \int_0^{\tau} e^{-rt}\left(\delta(A_t) - (1-\theta)C(A_t)\right) dt\right] + \\ &\quad E_x\left[1_{\tau>\tau(\underline{x})} \int_0^{\tau} e^{-rt}\left(\delta(A_t) - (1-\theta)C(A_t)\right) dt\right] \\ &\leq -\frac{\underline{c}}{r} E_x[\left(1 - e^{-r\tau}\right) 1_{\tau<\tau(\underline{x})}] + \\ &\quad E_x\left\{\left[-\frac{\underline{c}}{r}\left(1 - e^{-r\tau(\underline{x})}\right) + \xi(x, \underline{x})\Xi\right] 1_{\tau>\tau(\underline{x})}\right\}. \end{aligned}$$

Let $x^* > 0$ be the unique solution (in x) of $-\frac{\underline{c}}{r}\left(1 - e^{-r\tau(\underline{x})}\right) + \xi(x, \underline{x})\Xi = 0$. Since ξ is nondecreasing in x, we have for all $x \leq \tilde{x} = \underline{x} \wedge x^*$, $W(x, \tau) \leq -\frac{\underline{c}}{r} E[\left(1 - e^{-r\tau}\right) 1_{\tau<\tau(\tilde{x})}] \leq 0$, the optimum $W(x, \tau) = 0$ being reached for $\tau \equiv 0$. This claim proves that, starting from any level x and for any stopping time τ, the stopping time $\tau^- = \tau \wedge \tau(\tilde{x})$ is at least as good as τ. In other words, we can restrict ourselves, in our search for optimality, to the set of stopping times $\tilde{T} = \{\tau \text{ s.t. } \tau \leq \tau(\tilde{x})\}$. ■

Proof of Theorem 1. First, we prove the necessary conditions, then the sufficient conditions. The proof of the necessary conditions is based a series of lemmas:

LEMMA 2 *Under Conditions 1–3, f is continuously differentiable and f' is bounded* *and left and right differentiable. Moreover, f satisfies the following equations:*

$$\begin{aligned}\frac{1}{2}\sigma^2(x)f_l''(x)+\mu(x)f'(x)-rf(x)+\delta(x)-(1-\theta)C_l(x) &= 0\\ \frac{1}{2}\sigma^2(x)f_r''(x)+\mu(x)f'(x)-rf(x)+\delta(x)-(1-\theta)C(x) &= 0,\end{aligned}$$

where $f_l''(x)$ (resp. $f_r''(x)$) is the left (resp. right) derivative of f' at x, and $C_l(x)$ is the left limit of C at x.

Proof. First consider the case where C is continuous. Let $\phi(x)=\delta(x)-(1-\theta)C(x)$. From Condition 1, there exists a fundamental solution[29] $\zeta(x,v)$ such that

$$f(x)=\int_{\mathbb{R}}\zeta(x,v)\phi(v)dv$$

and

$$\frac{1}{2}\sigma^2(x)\frac{\partial^2\zeta}{\partial x^2}(x,v)+\mu(x)\frac{\partial\zeta}{\partial x}(x,v)-r\zeta(x,v)=0.$$

A straightforward differentiation then shows the result. When C is discontinuous, the second part of Condition 3 implies that there is a countably finite number of discontinuities. A limit argument by approximating, continuous functions then show the result. The boundedness of f' comes from the boundedness of $\frac{\partial\zeta}{\partial x}(x,v)$, proved in Friedman (1974). ■

COROLLARY 3 *W satisfies the following equations on $[A_B,\infty)$:*

$$\frac{1}{2}\sigma^2(x)W_l''(x)+\mu(x)W'(x)-rW(x)+\delta(x)-(1-\theta)C_l(x) = 0 \quad (25)$$

$$\frac{1}{2}\sigma^2(x)W_r''(x)+\mu(x)W'(x)-rW(x)+\delta(x)-(1-\theta)C(x) = 0, \quad (26)$$

[29]See Friedman (1974).

where $W_l''(x)$ (resp. $W_r''(x)$) is the left (resp. right) derivative of W' at x, and $C_l(x)$ is the left limit of C at x. In particular, W solves ODE (9) at any continuity point of C.

Proof. Straightforward, from Lemma 2 and (8). ■

COROLLARY 4 *W' is bounded on $[0, \infty)$*

Proof. Straightforward, from (8) and the fact that f' is bounded on $[0, \infty)$. ■

COROLLARY 5 *If a PSD obligation C satisfies Conditions 1–3, then $\tilde{W}(x, y)$ is continuously differentiable in both components, and $\frac{\partial \tilde{W}}{\partial x}$ is left and right differentiable in x.*

Proof. This comes directly from the above lemma and (8). ■

LEMMA 3 *If a PSD obligation C satisfies Conditions 1–3, then the optimal default boundary A_B verifies $\frac{\partial \tilde{W}}{\partial x}(A_B, A_B) = 0$.*

Proof. From (7) and Lemma 1, A_B satisfies $\frac{\partial \tilde{W}}{\partial A}(x, A_B) = 0$. Moreover, we have for any y, $\tilde{W}(y, y) = 0$ (since the firm defaults immediately). Differentiating this last equation and using the fact that $\frac{\partial \tilde{W}}{\partial x}(x, A_B) = 0$ yields $\frac{\partial \tilde{W}}{\partial x}(A_B, A_B) = 0$. ■

Combining (8), the above lemmas, and the fact that $W(x) = \tilde{W}(x, A_B)$ concludes the proof of all necessary conditions but one. It remains to show that $A_B \leq \bar{x}$.

This is trivial since, for $A_t > \underline{x}$, the cash flow rate exceeds the coupon rate, whence it is never optimal to default.

The verification of the sufficient conditions is similar to the proof of Proposition 2.1 in Duffie and Lando (2001). Define a stochastic process χ_t as

$$\chi_t = e^{-rt} W(A_t) + \int_0^t e^{-rs} \phi_s \, ds \ ,$$

where for $x > A_B$, $W(x)$ is the solution of the ODE that satisfies all the conditions listed in the theorem, and $W(x) = 0$ for $x \leq A_{\dot{B}}$. Since W is C^1, an application of Itô's formula leads to

$$d\chi_t = e^{-rt} d(A_t) \ dt + e^{-rt} W'(A_t) \, \sigma(A_t) dB_t, \tag{27}$$

where

$$d(x) \equiv \frac{1}{2} W''(x) \, \sigma^2(x) + W'(x) \, \mu(x) - rW(x) + \phi(x).$$

Since by assumption W' is bounded, the second term is a martingale, and since $E_x \left[\int_0^\infty \left(e^{-rt} W'(A_t) \sigma(A_t) \right)^2 dt \right] < \infty$, $\int_0^t e^{-rs} W'(A_s) \, \sigma A_s \, dB_s$ is a uniformly integrable martingale, which implies that $E_x \left[\int_0^\tau e^{-rs} W'(A_s) \, \sigma A_s \, dB_s \right] = 0$ for any stopping time τ. By assumption

$$\phi(A_B) \leq 0. \tag{28}$$

This inequality means that when the firm declares bankruptcy, its cash flow $\delta = (r - x) A_B$ is less than the coupon payment. It is easy to verify that the drift of χ_t is never positive: $d(x)$ vanishes for $x > A_B$ since W solves the ODE, and negative for $x < A_B$, because of the inequality (28) and $W(x) = 0$ for $x < A_B$. Because of the non-positive drift, for any stopping time $T \in \mathcal{T}$, $q_0 \geq E(\chi_T)$, meaning

$$W(A_0) \geq E\left[\int_0^T e^{-rs}\phi_s\, ds + e^{-rT} W(A_T)\right].$$

For the stopping time τ, we have

$$W(A_0) = E\left[\int_0^\tau e^{-rs}\phi_s\, ds\right] \geq E\left[\int_0^T e^{-rs}\phi_s\, ds\right],$$

where the inequality follows from non-negativity of W. Therefore, the stopping time τ maximizes the value of the equity. ■

Proof of Theorem 2. The proof is based on the following lemma:

LEMMA 4 *Let C and D be asset-based PSD satisfying Conditions 1–3, and $A_B^C \leq A_B^D$. If $h \equiv C-D$ is not constant on $[A_B^D, \infty)$ and changes sign at most once from positive to negative on $[A_B^D, \infty)$, then, $W_0^C(x) > W_0^D(x)$ for any starting asset level $x \in (A_B^C, \infty)$.*

Proof. Without loss of generality, we assume that the tax rate θ is zero. First, assume that $A_B^C = A_B^D = A_B$. Since h changes sign at most once from positive to negative on $[A_B, \infty)$, there exist constants A_1, A_2 verifying $A_B \leq A_1 \leq A_2$ and such that $h > 0$ for $A \in [A_B, A_1)$, $h = 0$ for $A \in (A_1, A_2)$, and $h < 0$ for $A \in (A_2, \infty)$.[30] We first consider the case where $A_1 = A_B$. Necessarily, $A_2 < \infty$, otherwise h would be constant on $[A_B, \infty)$. Thus, h vanishes on $[A_B, A_2)$ and is negative on (A_2, ∞). It is easy to verify that for any PSD C with initial asset level x and default boundary A_B,

[30]By convention [a,a) and (a,a) equal the empty set. The precise values at A_1 and A_2 are unimportant.

we have

$$U_0^C(x) = E_x\left[\int_0^{\tau(A_B)} e^{-rs}C(A_s)\,ds\right] + (A_B - \rho(A_B))\xi(A_0, A_B). \tag{29}$$

Since (A_2, ∞) has a positive Lebesgue measure, (29) implies that $U_0^D(x) > U_0^C(x)$ for all $x \in (A_B, \infty)$. Equation (6) then allows us to conclude that $W_0^C(x) > W_0^D(x)$ for all $x \in (A_B, \infty)$.

We now consider the case in which $A_1 > A_B$ or, equivalently, $h(A_B) > 0$. We first show that $W_0^C(x) > W_0^D(x)$ for all $x \in (A_B, A_1)$. From (25)) and (26), we have for $H(x) \equiv W_0^C(x) - W_0^D(x)$:

$$\frac{1}{2}H_l''(x)\sigma^2(x) + H'(x)\mu(x) - rH(x) - h_l(x) = 0 \tag{30}$$

$$\frac{1}{2}H_r''(x)\sigma^2(x) + H'(x)\mu(x) - rH(x) - h(x) = 0, \tag{31}$$

where $H_l''(x)$ (resp. $H_r''(x)$) is the left (resp. right) derivative of H' at x, and $h_l(x)$ is the left limit of h at x, which exists according to Condition 3 and Theorem 1. Also from Theorem 1, $W^i(A_B) = 0$ and $(W^i)'(A_B) = 0$ for $i = C, D$. Therefore, $H(A_B) = H'(A_B) = 0$. Since $h(A_B) > 0$, it follows from (31) that $H_r''(A_B) > 0$. This implies that $H'(x) > 0$ and $H(x) > 0$ in a right neighborhood of A_B. Precisely, there exists $\eta > 0$ such that $H'(x) > 0$ and $H(x) > 0$ for $x \in (A_B, A_B + \eta)$. We now prove by contradiction that $H'(x) > 0$ for all $x \leq A_1$. Let y denote the first time at which $H'(y) = 0$. Necessarily, $H(y) > 0$. From (30) and the fact that $h(y) \geq 0$ for $y \leq A_1$, it follows that $H_l''(y) > 0$, contradicting the fact that y is the first time at which $H'(y) = 0$. Therefore, $H'(x) > 0$ and $H(x) > 0$ on $(A_B, A_1]$. Last, we prove that $H(x) > 0$ on (A_1, ∞). By definition of W^C, W^D, and A_B,

$$W_0^C(x) = E_x^Q \left[\int_0^{\tau^*} q_t \left(\delta_t - C\left(A_t\right) \right) dt \right] \text{ and}$$
$$W_0^D(x) = E_x^Q \left[\int_0^{\tau^*} q_t \left(\delta_t - D\left(A_t\right) \right) dt \right],$$

where $q_t = e^{-rt}$, $\tau^* = \tau(A_B)$. Therefore,

$$H(x) = -E_x^Q \left[\int_0^{\tau^*} q_t h(A_t) dt \right].$$

This, combined with the fact that $\tau(A_1) < \tau(A_B) = \tau^*$ and $\int_0^{\tau^*} = \int_0^{\tau(A_1)} + \int_{\tau(A_1)}^{\tau^*}$, implies that

$$H(x) = -E_x^Q \left[\int_0^{\tau(A_1)} q_t h(A_t) \right] + E_x^Q(e^{-r\tau(A_h)}) H(A_1)$$

for all $x > A_1$. Since $h(\cdot)$ is non-positive on (A_1, ∞) and $H(A_1) > 0$, it follows that $H(x) > 0$ for all $x \in (A_B, \infty)$, which concludes the proof of the lemma when $A_B^C = A_B^D = A_B$. Now we consider the case where $A_B^C < A_B^D$. Then, $W_0^C(x) > 0$ and $W_0^D(x) = 0$ for $x \in (A_B^C, A_B^D]$, whence the claim holds trivially on this interval. The rest of the proof is identical to the first part for $x > A_B^D$. ■

From this lemma, we first conclude the proof of the theorem for asset-based PSD. We proceed by contradiction. We assume first that $A_B^C = A_B^D = A_B$. Then, the pair (C, D) satisfies the conditions of the lemma, which allows us to conclude that $W_0^C(x) > W_0^D(x)$ for all $x > A_B$. From (6), we conclude in particular that, for $x = A_0$, we have $U_0^C < U_0^D$, which contradicts the hypothesis of Theorem 2. We now assume that $A_B^C < A_B^D$. Then, we can lower the value of the interests paid by D

uniformly, proceeding by translation: we consider the PSD D_ε that pays the interest function $D_\varepsilon = D - \varepsilon$. Then, with the assumption that D is in the efficiency domain of its translation class (Condition 4), we have $U_0^{D_\varepsilon} < U_0^D = U_0^C$. On the other hand, since interest payments decrease as ε increases, there exists a constant $\varepsilon_0 > 0$ such that $A_B^{D_{\varepsilon_0+}} \leq A_B^C \leq A_B^{D_{\varepsilon_0-}}$. Moreover, since $h = C - D$ is non-increasing and not constant, so is $h_\varepsilon \equiv C - D_\varepsilon = C - D + \varepsilon$. In particular, h_ε is not constant and changes sign at most once. Since D satisfies Conditions 2 and 3, it is easy to verify that so does D_ε, for all $\varepsilon > 0$. Therefore, the pairs (C, D_ε) with ε in a left neighborhood of ε_0 satisfy the hypothesis of the lemma, which implies that $W_0^C(x) > W_0^{D_{\varepsilon_0}}(x)$[31] for any starting asset level $x \in (A_B^C, \infty)$. From (6), we conclude that $U_0^C < U_0^{D_\varepsilon}$ for any ε in a right neighborhood of ε_0, which contradicts the fact that $U_0^{D_\varepsilon} \leq U_0^D = U_0^C$ for all $\varepsilon > 0$. ■

Proof of Theorem 4. The proof is based on the proof of Theorem 2. In the case of ratings-based PSD obligations it is easy to see that Conditions 1–3 are automatically satisfied. We suppose first that $A_B^C = A_B^D$. This implies that $G(A_B^C) = G(A_B^D)$. From Lemma 4, $U_0^C > U_0^D$. This contradicts the fact that $U_0^C = U_0^D$. Now suppose that $A_B^C < A_B^D$. Take $\varepsilon > 0$ such that $A_B^C = A_B^{D_\varepsilon}$. Then, $G(A_B^C) = G(A_B^{D_\varepsilon})$, and Lemma 4 implies that $U_0^C < U_0^{D_\varepsilon}$. Condition 2, on the other hand, implies that $U_0^{D_\varepsilon} < U_0^D = U_0^C$, which leads to a contradiction. Therefore, $A_B^C > A_B^D$. Since $U_0^C = U_0^D$, the result then follows from (6). ■

[31]Here, we use the fact that $W_0^{D_\varepsilon}(x)$ is continuous in ε, which is an easy consequence of Corollary 5

Bibliography

[1] Acharya, V., Das, S., and Sundaram, R. (2002) "A Discrete-Time Approach to No-arbitrage Pricing of Credit Derivatives with Rating Transitions," *Financial Analysts Journal*, May-June, 28–44.

[2] Arnold, V.I. (1973) *Ordinary Differential Equations*, MIT Press, Cambridge, MA.

[3] Asquith, P., Beatty, A., and Weber, J. (2002) "Performance Pricing in Debt Contracts," Working Paper, Penn State University and MIT.

[4] Athey, S. (2002) "Monotone Comparative Statics Under Uncertainty," *Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 187–223.

[5] Aubin, J.-P., Ekeland, I. (1984) *Applied Nonlinear Analysis*, Wiley, New York, NY.

[6] Baumol, W.J., Quandt, R.E. (1964) "Rules of Thumb and Optimally Imperfect Decisions," *American Economic Review*, Vol. 54, No. 2, pp. 23–46.

[7] Beatty, P., Dichev, I., and Weber, J. (2002) "The Role and Characteristics of Accounting-based Performance Pricing in Private Debt Contracts," Working Paper, Michigan, Penn State and MIT.

[8] Berge, C. (1963) *Topological Spaces*, Oliver and Boyd, Edinburgh, UK. Reprinted by Dover Publications, Mineola, NY, in 1997.

[9] BERNHEIM, B.D., PELEG, B., AND WHINSTON, M.D. (1987) "Coalition Proof Nash Equilibria, I: Concepts," *Journal of Economic Theory*, Vol. 42, No. 1, pp. 1–12.

[10] BERNHEIM, B.D., WHINSTON, M.D. (1986) "Menu Auctions, Resource Allocations, and Economic Influence," *Quarterly Journal of Economics*, Vol. 101, No. 1, pp. 1–31.

[11] BERNSTEIN, F., FEDERGRUEN, A. (2005) "Decentralized Supply Chains with Competing Retailers Under Demand Uncertainty," *Management Science*, Forthcoming.

[12] BERTSEKAS, D.P. (1995) *Nonlinear Programming*, Athena Scientific, Belmont, MA.

[13] BLACK, F., COX, J. (1976) "Valuing Corporate Securities: Some Effects of Bond Identures Provisions," *Journal of Finance*, Vol. 31, No. 2, pp. 351–367.

[14] CACHON, G. (2003) "Supply Chain Coordination with Contracts," in: Graves S., de Kok, T. (eds.), *Handbooks in Operations Research and Management Science: Supply Chain Management*, North-Holland, Amsterdam, NL, pp. 229-339.

[15] DAS, S., TUFANO, P. (1996) "Pricing Credit Sensitive Debt when Interest Rates, Credit Rates and Credit Spreads are Stochastic," *Journal of Financial Engineering*, Vol. 5, No. 2, pp. 161–198.

[16] DIAMOND, D. (1991) "Debt Maturity Structure and Liquidity Risk," *Quarterly Journal of Economics*, Vol. 106, No. 3, pp. 709–737.

[17] DIXIT, A., GROSSMAN, G.M., AND HELPMAN, E. (1997) "Common Agency and Coordination: General Theory and Application to Government Policy Making," *Journal of Political Economy*, 105, pp. 753–769.

[18] DILIBERTO, S.P., STRAUS, E.G. (1951) "On the Approximation of a Function of Several Variables by the Sum of Functions of Fewer Variables," *Pacific Journal of Mathematics*, Vol. 1, pp. 195–210.

[19] Duffie, J.D. (2001) *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ, third ed.

[20] Duffie, J.D., Lando, D. (2001) "The Term Structure of Credit Spreads with Incomplete Accounting Information," *Econometrica*, Vol. 69, No. 3, pp. 633–664.

[21] Duffie, D., Singleton, K. (1999) "Modeling Term Structures of Defaultable Bonds," *Review of Financial Studies*, Vol. 12, No. 4, pp. 687–720.

[22] Edgeworth, F.Y. (1897) "The Pure Theory of Monopoly." Reprinted in: Edgeworth, F.Y. (1925) *Papers Relating to Political Economy*, Macmillan, London, UK.

[23] Fischer, E., Heinkel, R., and Zechner, J. (1989) "Dynamic Capital Structure Choice: Theory and Tests," *Journal of Finance*, Vol. 44, No. 1, pp. 19–40.

[24] Fitch IBCA, Duff & Phelps (2001) "The Operation and Evolution of Catastrophe-Linked Bonds," *Special Report.*

[25] Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., and Touzi, N. (1999) "An Application of Malliavin Calculus to Monte Carlo Methods in Finance," *Finance and Stochstics*, Vol. 3, No. 4, pp. 391–412.

[26] Fournié, E., Lasry, J.-M., Lebuchoux, J., and Lions, P.-L. (1999) "An Application of Malliavin Calculus to Monte Carlo Methods in Finance, II, " *Finance and Stochstics*, Vol. 5, No. 2, pp. 201–236.

[27] Friedman, A. (1975) *Stochastic Differential Equations and Their Applications*, Vol. 1, Academic Press, New York, NY.

[28] Goldstein, R., Ju, N., and Leland, H. (1998) "A Generalized Modigliani-Miller Model of Dynamic Capital Structure," Working Paper, Fisher College of Business and Haas School of Business.

[29] Granot, F., Veinott, A.F. (1985) "Substitutes, Complements and Ripples in Network Flows," *Mathematics of Operations Research*, Vol. 10, No. 3, pp. 471–497.

[30] GUEDES, J., OPLER, T. (1996) "The Determinants of the Maturity of Corporate Debt Issues," *Journal of Finance*, Vol. 51, No. 5, pp. 1809–1833.

[31] HARRISON, J. M. (1985) *Brownian Motion and Stochastic Flow Systems*, Wiley, New York, NY.

[32] HILLION, P., VERMAELEN, T. (2004) "Death Spiral Convertibles," *Journal of Financial Economics*, Vol. 71, No. 2, pp. 381–415.

[33] HIRSCH, F. (1976) *Social Limits to Growth*, Harvard University Press, Cambridge, MA.

[34] HOUWELING, P., MENTINK, A., AND VORST, T. (2003) "Valuing Euro Rating-Triggered Step-Up Telecom Bonds," Working Paper, Erasmus University Rotterdam.

[35] JARROW, R., LANDO, D., AND TURNBULL, S. (1997) "A Markov Model for the Term Structure of Credit Risk Spreads," *Review of Financial Studies*, Vol. 10, No. 2, pp. 481–523.

[36] JARROW, R., TURNBULL, S. (1995) "Pricing Derivatives on Financial Securities Subject to Credit Risk," *Journal of Finance*, Vol. 50, No. 1, pp. 53–86.

[37] KAKUTANI, S. (1941) "A Generalization of Brouwer's Fixed Point Theorem," *Duke Mathematical Journal*, Vol. 8, pp. 457–459.

[38] KARATZAS, I. (1984) "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes," *Annals of Probability*, Vol. 12, No. 1, pp. 173–192.

[39] KARATZAS, I., SHREVE, S.E. (1991) *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, NY.

[40] KHALIL, H.K. (1992) *Nonlinear Systems*, Macmillan, New York, NY.

[41] KHAVINSON, S.YA. (1997) *Best Approximation by Linear Superpositions (Approximate Nomography)*, American Mathematical Society, Providence, RI.

[42] KIM, I., RAMASWAMY, K., AND SUNDARESAN, S. (1993) "Valuation of Corporate Fixed-Income Securities," *Financial Management*, Special Issue on Financial Distress, pp. 117 – 131.

[43] KOLMOGOROV, A.N. (1957) "On the Representation of Continuous Functions of Several Variables by Superpositions of Continuous Functions of One Variable and Addition," *Doklady Akademii Nauk SSSR*, Vol. 108, pp. 953–956. English translation in: *American Mathematical Society Translations* (1961), No. 28, pp. 369–373.

[44] LANDO, D., MORTENSEN, A. (2003) "Mispricing of Step-Up bonds in the European Telecom Sector," Working Paper, Copenhagen Business School.

[45] LEE, J.M. (2003) *Introduction to Smooth Manifolds*, Springer, New York, NY.

[46] LELAND, H. (1994) "Corporate Debt Value, Bond Covenants, and Optimal Capital Structure," *Journal of Finance*, Vol. 49, No. 4, pp. 1213–1252.

[47] LELAND, H. (1998) "Agency Costs, Risk Management, and Capital Structure," *Journal of Finance*, Vol. 53, No. 4, pp. 1213–1243.

[48] LELAND, H., TOFT, K. (1996) "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads," *Journal of Finance*, Vol. 51, No. 3, pp. 987–1019.

[49] LONGSTAFF, F., SCHWARTZ, E. (1995) "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt," *Journal of Finance*, Vol. 50, No. 3, pp. 789–820.

[50] MAGARIL-IL'YAEV, G.G., TIKHOMIROV, V.M. (2003) *Convex Analysis: Theory and Applications*, American Mathematical Society, Providence, RI.

[51] MAS-COLELL, A. WHINSTON, M.D., AND GREEN, J.R. (1995) *Microeconomic Theory*, Oxford University Press, New York, NY.

[52] MATTHEWS, S., MOORE, J. (1987) "Monopoly Provision of Quality and Warranties: An Exploration of the Theory of Multidimensional Screening," *Econometrica*, Vol. 55, No. 2, pp. 441–467.

[53] MATTHEWS, S.A. (1995) "Renegotiation of Sales Contracts," *Econometrica*, Vol. 63, No. 3, pp. 567–589.

[54] MERTON, R.C. (1974) "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, Vol. 29, No. 2, pp. 449–470.

[55] MILGROM, P.R. (1994) "Comparing Optima: Do Simplifying Assumptions Affect Conclusions?", *Journal of Political Economy*, Vol. 102, No. 3, pp. 607–615.

[56] MILGROM, P.R., ROBERTS, J. (1990) "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, Vol. 80, No. 3, pp. 511–528.

[57] MILGROM, P.R., ROBERTS, J. (1994) "Comparing Equilibria," *American Economic Review*, Vol. 84, No. 3, pp. 441–459.

[58] MILGROM, P.R., SHANNON, C. (1994) "Monotone Comparative Statics," *Econometrica*, Vol. 62, No. 1, pp. 157–180.

[59] MIRMAN, L.J., SIBLEY, D.S. (1980) "Optimal Nonlinear Prices for Multiproduct Monopolies," *Bell Journal of Economics*, Vol. 11, No. 2, pp. 659–670.

[60] MIRRLEES, J.A. (1971) "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, Vol. 38, No. 2, pp. 175–208.

[61] MOODY'S (2001) "The Unintended Consequences of Rating Triggers," *Special Comment.*

[62] MOODY'S (2002) "Moody's Analysis of US Corporate Rating Triggers Heightens Need for Increased Disclosure," *Special Comment.*

[63] MUSSA, M., ROSEN, S. (1978) "Monopoly and Product Quality," *Journal of Economic Theory*, Vol. 18, No. 2, pp. 301–317.

[64] MYERS, S. (1977) "Determinants of Corporate Borrowings," *Journal of Financial Economics*, Vol. 5, No. 1, pp. 147–175.

[65] NUALART, D. (1995) *The Malliavin Calculus and Related Topics*, Springer, New York, NY.

[66] PRAT, A., RUSTICHINI, A. (2003) "Games Played Through Agents," *Econometrica*, Vol. 71, No. 4, pp. 989–1026.

[67] ROBERTS, K.W.S. (1979) "Welfare Considerations of Nonlinear Pricing," *Economic Journal*, Vol 89, No. 353, pp. 66–83.

[68] ROCHET, J.-C., STOLE, L.A. (2003) "The Economics of Multidimensional Screening," in: Dewatripont, M., Hansen, L.-P., Turnovsky, S.J. (Eds.), *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress*, Vol. I, Cambridge University Press, New York, NY, pp. 150–197.

[69] SAMUELSON, P.A. (1941) "The Stability of Equilibrium: Comparative Statics and Dynamics," *Econometrica*, Vol. 9, No. 2, pp. 97–120.

[70] SAMUELSON, P.A. (1974) "Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory," *Journal of Economic Literature*, Vol. 12, No. 4, pp. 1255–1289.

[71] SEGAL, I. (1999) "Contracting With Externalities," *Quarterly Journal of Economics*, Vol. 104, No. 2, pp. 337–388.

[72] SEGAL, I., WHINSTON, M.D. (2003) "Robust Predictions for Bilateral Contracting with Externalities," *Econometrica*, Vol. 71, No. 3, pp. 757–791.

[73] STANDARD & POOR'S (2001) "Playing Out the Credit Cliff Dynamic," *Standard & Poor's Ratings Criteria.*

[74] TOPKIS, D.M. (1968) *Ordered Optimal Solutions*, Doctoral Dissertation, Stanford University, Stanford, CA.

[75] TOPKIS, D.M. (1998) *Supermodularity and Complementarity*, Princeton University Press, Princeton, NJ.

[76] VEBLEN, T. (1899) *The Theory of the Leisure Class.* Reprinted by Penguin Books, New York, NY, in 1994.

[77] VEINOTT, A. F. JR. (1965) "Optimal Policy for a Multi-Product, Dynamic, Nonstationary Inventory Problem," *Management Science*, Vol. 12, No. 3, pp. 206–222.

[78] WEBER, T.A., XIONG, H. (2004) "Efficient Contract Design in Multi-Principal Multi-Agent Supply Chains," Working Paper, Department of Management Science and Engineering, Stanford University, Stanford, CA.

[79] ZORICH, V.A. (2004) *Mathematical Analysis, Volume I+II*, Springer, New York, NY.