Essays on Privatization, Identity, and Political Polarization

Erik Lindqvist

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics

KEYWORDS: Privatization, Service contracting, Youth care, Identity, Ethnicity, Social Class, Redistribution, Political Polarization, Economic Performance

ISBN NR 978-91-7258-733-5

PRINTED BY:
Elanders Gotab, Stockholm 2007

DISTRIBUTED BY:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm
www.hhs.se/efi

To Karin

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization

The research activities at the Institute are organized into 21 Research Centres.
Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	*Centre Director:*
Management and Organisation (A)	Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr (acting)
Centre for Innovation and Operations Management (T)	Christer Karlsson
Centre for Media and Economic Psychology (P)	Guje Sevón
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Centre for Business and Economic History (BEH)	Håkan Lindgren
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Bengt Jönsson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Paul Segerstrom
Economic Statistics (ES)	Anders Westlund
Law (RV)	Johnny Herre
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address

EFI, Box 6501, SE-113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Contents

Acknowledgements

Coming from the bohemian existence of undergraduate studies, life as a graduate student has a lot to offer. You get a desk of your own, an empty whiteboard, and direct access to a coffee machine. You spend your time doing more or less what you want, and, to my perpetual surprise, you get paid for it. On Tuesdays you eat cake. On Fridays you eat cake again, and then there is Friday Beer. The most important component of life as a graduate student, however, is the people you interact with. Now is the time to thank those people.

First, I want to express my sincere gratitude to my advisor, Tore Ellingsen. Tore accepted being my advisor despite a hectic work schedule, a choice at least *I* have never regretted. Tore has been very supportive throughout the work with this dissertation, with a great ability to instill enthusiasm in research and make obstacles seem surmountable.

My co-author and roommate Robert Östling has made my graduate studies a lot easier, funnier, and more interesting. Among other things, Robert has provided computer wizardry, T-shirts in funny colors and a malfunctioning toy railway to our office.

The paper on residential youth care could not have been written unless Bo Vinnerljung and Marie Sallnäs at the Institute for Evidence-Based Social Work Practice (IMS) had very generously provided me with their data on youth care. During the work with this paper, I have come to appreciate the richness of this data set, and the vast amount of work required to compile it. I have also relied heavily on the work by Marie Sallnäs regarding the institutional details of Swedish youth care.

The papers on privatization both stem from my final thesis at the undergraduate program at Stockholm School of Economics, which I wrote together with Jens Sandblom. The basic ideas of those papers were already in place when Jens and I wrote our thesis, and Jens has made a significant contribution to both papers.

I benefited a lot from the academic year I spent at the Department of Economics at the Massachusetts Institute of Technology, and I am thankful for their hospitality.

I would also like to thank my fellow graduate students at the Stockholm School of Economics and elsewhere for friendship, help and interesting discussions: Johan Almenberg, Milo Bianchi, Anna Breman, David Cesarini, Anna Dreber, Max Elger,

Erika Färnstrand Damsgaard, Ola Granström, Erik Höglin, Tobias Lundquist, Henrik Lundvall, Kristin Magnusson, Ola Melander, Damian Migueles, Erik Mohlin, Martin Mossberg, Eva Ranehill, Marcus Salomonsson, Per Sonnerby, Björn Tyrefors, Björn Wallace and Robertas Zubrickas. A special thanks to my cohabitants in Somerville, Massachusetts: Anders Åkerman, Martin Bech Holte and Sebastian Giwa; and to Roine Vestman.

The administrative staff at the Department: Carin Blanksvärd, Ritva Kiviharju, Anneli Sandbladh and Lillian Öberg, have been very friendly and helpful. I have enjoyed working as a TA on Richard Friberg's micro course a lot, and benefited from discussions with Justina A. V. Fischer, Magnus Johannesson, Giancarlo Spagnolo and Jörgen Weibull.

I am thankful for financial support from the Tom Hedelius and Jan Wallander Foundation, which has financed my two last years as a graduate student, including my year at Massachusetts Institute of Technology.

My parents have kept me well-nourished, and have repeatedly reminded me of the value of vacation. My aunt, Karin Borgström, has still not got back a couple of books on youth care she lent to me five years ago. My sister, Ylva, have told me stories from the *real* world of social work.

My final thanks goes to my girlfriend Karin, a living proof that size and significance are two quite distinct matters.

Stockholm, September 2007

Erik Lindqvist

Summary of Thesis

Introduction

This dissertation consists of two papers on service contracting in the public sector (also referred to as "privatization"), and two papers on political economics.

The papers on privatization are the result of a project I started as an undergraduate in 2002 together with Jens Sandblom. The first paper discusses service contracting in general, with no focus on any particular service. The second paper focuses on the case of residential youth care. Though privatization is a hotly debated issue, the purpose of this research is to better understand its consequences, not to argue for any particular political ideology.

The first paper on political economics discusses the interaction between identity and redistribution, and the second paper how political polarization may affect economic and government performance. Though these papers are only weakly connected to each other, an argument common to both papers is that people's extent of altruism (which in turn affects redistribution), is affected by the perceived difference to other people in society.

Summary of Papers

Paper 1: Will Privatization Reduce Costs?

Governments often procure services, such as garbage collection or elderly care, from private firms. A central issue in the debate on the merits of privatization has been how private and public firms differ regarding quality and cost. The standard view of economic theory is that private firms are cheaper than public firms, but that they might shirk on quality provision (Schmidt 1996, Hart, Shleifer and Vishny 1997, Shleifer 1998).

In this paper, I develop a model of public sector contracting based on the multitask framework by Holmström and Milgrom (1991). In the model, an agent can put effort into increasing the quality of a service or reducing costs. Being residual claimants, private owners have stronger incentives to cut costs than public employees. However, if quality cannot be perfectly measured, providing a private firm with incentives to improve quality forces the owner of the firm to bear risk. As a result, private firms will always be cheaper for low levels of quality but might be more expensive for high levels of quality. Allowing for contracting on cost, I find that although contracts are complete in the sense that a private firm can be provided with exactly the same incentives as a public firm, ownership still matters. If effort on quality and cost are sufficiently hard to measure, public provision is still cheaper than private provision for high levels of quality. Extending the model to allow for differences in task attractiveness, I find that public firms shun unattractive tasks, whereas private firms undertake them if incentives are strong enough.

Paper 2: When Does Privatization Reduce Costs and Improve Quality? Theory and Evidence from Service Contracting

In this paper, I develop and test a model of service contracting in residential youth care. Being residual claimants, private providers have stronger incentives to keep teenagers in care and cut costs, whereas managers of public facilities lack financial incentives. This difference generates three main predictions that differ from the standard view in the privatization literature. First, when the duration of treatment is

insensitive to quality, owners of private facilities must be given rents in order to invest in quality. As a result, private provision is more expensive for high levels of quality, though still cheaper for low levels of quality. Second, as public sector managers lack financial incentives, they shun troublesome teenagers. Third, private providers try to prolong the duration of treatment by lying about treatment progress, whereas managers in public facilities tell the truth.

The model's predictions are tested on a rich data set of Swedish teenagers placed in youth care during 1991. Using personnel density as a quality measure, I find that private facilities have lower per-day cost of treatment for low levels of quality, but higher per-day cost for high levels of quality. Accordingly, private facilities have lower personnel density, a comparably low level of education among the personnel and a higher frequency of treatment breakdowns. However, as predicted by theory, the risk of a treatment breakdown for teenagers with severe problems is higher in public facilities than in private facilities. While violent and non-violent teenagers experience almost the same probability of treatment breakdown in private facilities, the breakdown frequency is three times higher for violent teenagers than for non-violent teenagers in public facilities. In line with the model's third prediction, the average duration of treatment in private facilities is twice that in public facilities. As the average per-day cost of treatment is only ten percent lower in private facilities, this implies that the total cost of treatment under private provision is almost double that of public provision.

I use data on post-treatment outcomes at the age of 25 to evaluate the long-term effects of facility ownership. For non-troublesome teenagers, there are no significant differences in terms of criminal record, mental health, economic self-reliance and educational attainment between private and public facilities. This suggests that the longer duration of treatment in private facilities compensates for the lower level of quality. However, private facilities are relatively better at treating troublesome teenagers. Violent teenagers in private facilities are relatively less likely to have been convicted in a court of law or to have been imprisoned between the age of 20 and 24 compared to violent teenagers in public facilities. As total cost is twice as high under private provision, public provision of youth care appears to be superior for non-troublesome teenagers, but it is hard to make a definitive assessment for troublesome teenagers.

Finally, the prevalence of treatment breakdowns and recidivism are lower when the municipality social service plan the duration of treatment in advance, regardless of facility ownership. This suggests that the social services engagement is important for youth care to benefit teenagers.

Paper 3: Identity and Redistribution (co-author Robert Östling)

Both the canonical economic theory of redistribution (e.g. Meltzer and Richard 1981) and Marxian theory assume that people's political preferences are determined by their economic position in society. This view is controversial. Conflicts along other dimensions, such as ethnicity, race, religion or gender, may be more important than social class.

The view that there are multiple dimensions of political conflict invokes the questions under what circumstances voters come to identify with a certain group and how their identities affect redistributive policies. Social psychology research indicates that people tend to identify with groups that have high status, suggesting that redistribution in turn might affect identity choices.

In this paper, we model the interaction between individuals' identity choices and redistribution. Both redistributive polices and identity choices are endogenous, and there might be multiple equilibria. The model is applied to ethnicity and social class. In an equilibrium with high taxes, the poor identify as poor and favor high taxes. In an equilibrium with low taxes, at least some of the poor identify with their ethnic group and favor low taxes. The model has two main predictions. First, redistribution is highest when society is ethnically homogenous, but the effect of ethnic diversity on redistribution is not necessarily monotonic. Second, when income inequality is low, an increase in income inequality might induce the poor to identify with their ethnic group and therefore favor lower taxes.

Paper 4: Political Polarization and Economic Performance (co-author Robert Östling)

It is commonly believed that political polarization of the population has detrimental effects on economic performance. For example, the European Council has adopted a policy for "minimizing disparities and avoiding polarization", partly because of its presumed effect on economic development. Economic theory suggests that political polarization may reduce the support for spending on public goods (Alesina et al., 1999), reduce redistribution to the poor (Fernández and Levy, 2007), and increase the scope for rent-seeking if politicians are less constrained by a polarized population (Persson and Tabellini, 2000). Despite the mechanisms suggested by theory and the frequency of appeals to unity by political leaders, there are few empirical studies on the economic effects of political polarization among the population. Moreover, most studies that do consider polarization among the population use indirect measures of political preferences, such as ethnicity or income.

In this paper, we use responses to multiple choice questions about economic policy to derive measures of political polarization at the country level. Our main measure is the standard deviation of responses, but we also consider several alternative measures. All these measures are strong predictors of economic and government performance. Countries that are more cohesive have higher income per capita, more efficient and less corrupt governments, a higher output of public goods and more redistribution of income.

Although we are not able to make a definite test of the direction of causality, there is some evidence suggesting that these relationships reflect a causal effect of political polarization on economic and government performance. First, the relationship between polarization and performance persists when we use ethnic fractionalization and polarization in political interest as instruments for political polarization. Second, the levels of political polarization in West and East Germany in 1990 – the year of the reunification – are remarkably similar, indicating that polarization is stable over time and not endogenous to economic performance. In addition, polarization of political preferences does not seem to be the result of income inequality per se, divisive economic policies, an unwillingness to give extreme answers in certain countries, or higher uncertainty about how to answer survey questions in developing countries. Further, the relationships between polarization of preferences and economic outcomes seem to be stronger in democracies, indicating that part of the effect goes through the political system.

References

Alesina, A., Baqir, R. and Easterly, W. (1999), "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics* 115(4), 1167–1199.

Fernández, R. and Levy, G. (2007), "Diversity and Redistribution", NBER Working Paper 11570.

Hart, O., Shleifer, A. and Vishny, R. W. (1997), "The Proper Scope of Government: Theory and an Application to Prisons", *Quartely Journal of Economics*, November 1997, Vol. 112, 1127–1161.

Holmström, B. and Milgrom, P. (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics and Organization*, Vol 7, 24–51.

Meltzer, A. H. M. and Richard, S. F. (1981), 'A Rational Theory of the Size of Govern ment', *Journal of Political Economy* 89(5), 914–927.

Persson, T. and Tabellini, G. (2000), *Political Economics: Explaining Economic Policy*, The MIT Press, Cambridge, MA.

Schmidt, K. M. (1996), "The Costs and Benefits of Privatization: An Incomplete Contracts Approach", *Journal of Law, Economics and Organization*, Vol. 12, No. 1, 1–24.

Shleifer, A. (1998), "State versus Private Ownership", *Journal of Economic Perspectives*, Fall 1998.

Papers

PAPER 1

Will Privatization Reduce Costs?

Erik Lindqvist

ABSTRACT. I develop a model of public sector contracting based on the multitask framework by Holmström and Milgrom (1991). In this model, an agent can put effort into increasing the quality of a service or reducing costs. Being residual claimants, private owners have stronger incentives to cut costs than public employees. However, if quality cannot be perfectly measured, providing a private firm with incentives to improve quality forces the owner of the firm to bear risk. As a result, private firms will always be cheaper for low levels of quality but might be more expensive for high levels of quality. Extending the model to allow for differences in task attractiveness, I find that public firms shun unattractive tasks, whereas private firms undertake them if incentives are strong enough.

1. Introduction

Governments often procure services, such as garbage collection or elderly care, from private firms. A central issue in the debate on the merits of privatization has been how private and public firms differ regarding quality and cost. The standard view of economic theory is that private firms are cheaper than public firms, but that they might shirk on quality provision (Schmidt 1996, Hart, Shleifer and Vishny 1997, Shleifer 1998). In this paper, I develop a model of privatization that combines the multitask framework by Holmström and Milgrom (1991) with incomplete contracts.[1] My main result is that private firms are cheaper for low levels of quality, but that they might be more expensive for high levels of quality when quality is measured imperfectly.

[0] This work originates from an unpublished paper co-authored by Jens Sandblom, to whom I am very grateful. I am also indebted to Tore Ellingsen. In addition, I would like to thank Martin Bech Holte, Daniel Ferreira, Guido Friebel, Bengt Holmström, Jonathan Levin, Paul Milgrom, Robert Östling, Per Sonnerby and seminar participants at the Econometric Society European Meeting 2006, Massachussetts Institute of Technology, Stockholm Institute of Transition Economics and the Research Institute of Industrial Economics (Stockholm) for valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[1] The model could also be thought of as a general model of contracting out, applicable to the make-or-buy decision of private firms. Yet in this paper we will restrict our discussion to the case of privatization.

As shown by Sappington and Stiglitz (1987), ownership matters only if contracts are incomplete. If it were possible to write complete contracts, taking every contingency over the entire horizon of the firm into account, the government could provide a private and a public firm with exactly the same incentives. Many models of incomplete contracting build on Grossman and Hart (1986), who assume that ownership allocates residual control rights that influence bargaining power in later stages of a contractual relationship. Notably, Hart, Shleifer and Vishny (1997) assume that an owner of a private firm has a stronger bargaining position toward the government than a manager of a public firm. Therefore, the private provider has stronger incentives to implement innovations that improve quality or reduce costs. However, the private provider's incentives to cut costs might be too strong since he ignores any negative effects on quality that the cost reduction might entail. As a result, private provision will always be cheaper than public provision, but may be either superior or inferior in terms of quality.

A distinguishing feature of Hart, Shleifer and Vishny (1997) is that ex ante contracts are the same under private and public provision. This view contrasts with the extensive use of procurement agencies in developed economies. Whereas explicit contracts are scarce under public provision, privatization in service contracting seems to go hand in hand with extensive ex ante contracting. In this paper, I take the view that there is some scope for ex ante contracting on quality with private providers, but not with public providers. The basic difference between the view of Grossman and Hart (1986) and this paper is that whereas they distinguish between the ownership of assets and contractual compensation, I assume that integration shrinks the set of feasible compensation contracts. If the government holds the residual control rights, contracts on profit sharing or rewards for the completion of tasks could be manipulated ex post. Hence, like Schmidt (1996), I view privatization as a commitment device of the government that resolves a time-inconsistency problem.

The agent in my model can invest effort in improving service quality or reducing costs. Because of the commitment problem, employees of public firms cannot be given strong incentives to engage in either of these tasks, implying that work effort is low but balanced between quality improvements and cost reductions. In contrast, owners of private firms always have strong incentives to cut costs and must be provided with financial incentives for quality provision in order to improve quality. As a result, when only weak incentives are feasible, the maximum level of quality is higher under public provision than under private provision. On the other hand, if strong incentives are feasible, the maximum level of quality is higher under private provision.

Although it is possible to contract on quality outcomes with a private provider, outcomes may only partly reflect effort. When this is the case, providing a private firm with incentives to improve quality will force the owner of the firm to bear risk, raising the price for a given level of quality. Public employees are insensitive to such measurement problems since they are paid a fixed wage regardless of the outcome. If owners of private firms are sufficiently risk averse, private firms are more expensive than public firms for high levels of quality, even though they operate more efficiently. I thus identify a potential non-monotonicity regarding the choice of ownership structure in government preference for quality: If strong incentives for quality are feasible, but effort on quality is hard to measure, private provision is preferable when quality is considered either unimportant or very important whereas public provision is optimal when quality is of intermediate importance. This result differs from the prediction of Hart, Shleifer and Vishny (1997) that private provision is always cheaper.

The model formalizes arguments that have previously been put forward in an informal way. For example, Donahue (1989, p. 87 and p. 105) and Sclar (2000, p. 109) both argue that the inability to accurately measure effort on quality may lead to an suboptimal allocation of risk under private provision.

I present the model in the next section, and the main results are laid out in Section 3. In Section 4, I discuss some extensions of the basic model. Allowing for contracting on cost, I find that although contracts are complete in the sense that a private firm can be provided with exactly the same incentives as a public firm, ownership still matters. If effort on quality and cost are sufficiently hard to measure, public provision is still cheaper than private provision for high levels of quality. Extending the model to allow for differences in task attractiveness, I find that public firms shun unattractive tasks, whereas private firms undertake them if incentives are strong enough. Section 5 concludes the paper.

2. The Model

The principal is a public agency that has to decide whether to procure a service from the market or produce it in-house. The agent is a service provider. When service provision is organized in-house, I refer to the service provider as the manager, with private provision, the service provider is called owner.

A service provider can put effort on to two different tasks. The first task (t_q) improves the quality of the service and the other (t_c) reduces its costs. Engagement in the two tasks entails a cost $C(t)$ for the service provider, where $t = t_q + t_c$. The cost function, $C(t)$, is strictly convex, twice continuously differentiable and minimized at

$t^* > 0$, i.e., people are assumed to exert some effort even in the absence of financial incentives.

The outcome of the cost reducing task is certain, but effort on cost reductions is not observable to the public agency.[2] The cost reduction is given by a function $S(t_c)$ which is increasing, strictly concave and twice continuously differentiable in t_c. Total cost is thus decreasing in $S(t_c)$. In contrast to cost savings, the effort on quality is observable, albeit imperfectly. Formally, the public agency receives a signal of quality $q = t_q + \varepsilon_q$, where ε_q is a stochastic term with distribution $\varepsilon_q \sim N\left(0, \sigma_q^2\right)$. It is only possible to contract on outcomes (q) and not directly on effort (t_q).

The public agency's valuation of service quality is described by a function $B\left(t_q, \psi\right)$, which is increasing and strictly concave in t_q for all $\psi > 0$. ψ is a parameter that denotes the public agency's valuation of quality. I assume that $B\left(t_q, \psi\right)$ is twice continuously differentiable in both t_q and ψ, $\lim_{\psi \longrightarrow \infty} B\left(t_q, \psi\right) = \infty$, $B_{t_q}\left(t_q, 0\right) = 0$ for all t_q and $\lim_{\psi \longrightarrow \infty} B_{t_q}\left(t_q, \psi\right) = \infty$ for all t_q, where $B_{t_q}\left(t_q, \psi\right)$ denotes the partial derivative of $B\left(t_q, \psi\right)$ with respect to t_q. The service provider's utility is defined as $u\left(x\right) = -e^{-rx}$, where r is a measure of risk aversion and x is the financial return minus the cost of effort. I assume that $r > 0$. The public agency is risk neutral.

The service provider receives a pecuniary compensation $w\left(q\right) = \alpha q + \beta$ and a share of cost savings $\lambda \in [0, 1]$. The parameter α is thus a measure of the economic incentive to engage in the quality improving task and β is a fixed wage component. Similarly, λ denotes the incentive to reduce costs. We get $x = w\left(q\right) + \lambda S(t_c) - C\left(t\right)$ and the manager/owner's expected utility is thus

$$E\left[u\left(w\left(q\right) - C\left(t\right) + \lambda S\left(t_c\right)\right)\right] = u\left(CE\right),$$

where CE is the certainty equivalent. Solving for CE, we get

$$CE = \alpha t_q + \beta - C\left(t\right) + \lambda S(t_c) - \frac{1}{2} r \sigma_q^2 \alpha^2,$$

where $\frac{1}{2} r \sigma_q^2 \alpha^2$ is the agent's risk premium. I refer to the product $r\sigma_q^2$ as the cost of risk.

The agent's outside option is normalized to zero, implying that the participation constraint is $CE \geq 0$. Assuming perfect competition, the fixed wage component β will be set such that the participation constraint always binds.

How do private and public firms differ in this model? The key assumption is that the public agency cannot credibly commit to high-powered incentives, e.g., payment depending on the completion of tasks or profit sharing, under public (in-house) provision. In terms of the model, this means that $\alpha = 0$ and $\lambda = 0$. The argument

[2] I relax the assumption that effort on cost reductions is not contractible in Section 4.1.

for this assumption is that ownership gives control over key contractual parameters, which enlarges the public agency's scope for opportunistic behavior. For example, if the manager has a share in the firm's profits, the owner could influence stated profits by changing accounting procedures. Alternatively, the public agency could override the manager's decisions in a way that inflicts on his chances of reaching performance clauses. Another fundamental commitment problem is the existence of the public firm itself. If the public agency cannot commit to not sell or close down the public firm in the future, it is hard to commit to incentives for outcomes that are observable only in the long run. This in turn affects the optimal level of incentives for tasks that are observable in the short run as strong incentives on such tasks will crowd out effort on tasks where outcomes are only observable in the long run. Hence, the inability to commit to long-run incentives creates a problem of intertemporal incentive balance. In practice, there are also some less fundamental reasons for why it might be difficult to provide managers of public firms with strong incentives. In many countries, employment protection laws and union wage bargaining restrict the space of feasible employment contracts. For example, it might more difficult to fire an employee should he deviate from the contractual terms than cease business with a private firm.

Private firms have a stronger legal position toward the public agency in the sense that it is harder for the public agency to manipulate contracts ex post. In addition, contracts with private firms are not constrained by employment regulations. However, even though contractual compensation under private provision can be based on observable outcomes, not every contract is credible ex ante. For example, if bankruptcy of the private provider implies that service delivery is interrupted, a contract that entails bankruptcy for the private firm in some states of the world might not be renegotiation-proof. Moreover, public procurement procedures are often subject to regulations that aim to reduce the risk of collusion between bureaucrats and firms. As pointed out by Wilson (2000, p. x), such regulations make it hard for the government purchaser to take past experience of a firm into account, thereby limiting the scope for implicit contracting on quality. If high-powered incentives for quality provision are not credible ex ante, the set of possible contracts is limited by some upper bound on α, i.e. $\alpha \in [0, \overline{\alpha}]$.[3]

Since effort on the cost reducing task is unobservable to the public agency and the owner of the private firm has the residual control rights, $\lambda = 1$ in case of private provision. The assumption that cost savings are unobservable might seem stark and I will relax it in an extension below. However, the assumption makes more sense if we include activities that affect the firm's cost structure over the long term in the definition of cost

[3] We can think of α as being the result of both explicit and implicit contracting (such as relational contracting or reputation effects), though the game played by principal and agent is not modelled explicitly.

savings. In this case, a wide range of activities that might have a negative immediate impact on profits would count as effort on cost reductions. Examples include many tasks which arguably are very hard to contract on directly, such as equipment maintenance, the development of administrative systems and human resource management.

2.1. Public firm. Since the public agency cannot commit to incentive payment schemes under public provision, α and λ are equal to zero. The manager's maximization problem is therefore[4]

$$\max_{t_q,t_c} \quad \beta - C(t). \tag{2.1}$$

Since the agent is indifferent between the two tasks he just minimizes his cost function with respect to total effort.[5] This gives the optimality condition $t_q + t_c = t^*$ where $C'(t^*) = 0$. Since the manager is indifferent between tasks I assume that he divides his effort according to the wishes of the public agency.[6] The public agency's maximization problem is then

$$\begin{array}{ll} \max\limits_{t_q,t_c} & B(t_q,\psi) + S(t_c) \\ \text{s.t.} & t_q + t_c = t^*. \end{array} \tag{2.2}$$

Let the superscript m denote the manager of the public firm. The solution to this problem is

$$\left(t_q^m, t_c^m\right) = \begin{cases} t_q = 0 \text{ and } t_c = t^* & \text{if } B_{t_q}(0,\psi) \leq S'(t^*); \\ \left\{(t_q,t_c) |\, B_{t_q}(t_q,\psi) = S'(t_c)\right\} & \text{if } B_{t_q}(0,\psi) > S'(t^*) \text{ and } B_{t_q}(t^*,\psi) < S'(0); \\ t_q = t^* \text{ and } t_c = 0 & \text{if } B_{t_q}(t^*,\psi) \geq S'(0). \end{cases}$$

Hence, the manager of the public firm sets the total level of effort and the public agency decides how this effort should be divided between the two tasks. The public agency orders the manager to set $t_q = 0$ if the marginal value of quality at $t_q = 0$ is weakly lower than the marginal benefit of cost savings at $t_c = t^*$. When the public agency's valuation of quality is higher, the manager is instructed to put effort on quality until the marginal value of quality equals the marginal value of cost reductions. The second corner solution occurs when the public agency's marginal valuation of quality at $t_q = t^*$ is weakly higher than the marginal returns to cost savings at $t_c = 0$. The division of effort between tasks is therefore optimal under public provision, whereas total work effort is suboptimally low.

---

[4] Note that since $u(CE)$ is a monotone transformation of CE, maximizing CE is equivalent to maximizing $u(CE) = E[u]$.

[5] The assumption that the manager is indifferent between tasks is discuseed in Section 4.2.

[6] I thus implicitly assume that the agent has lexicographic preferences where his first priority is to maximize his own utility and his second priority to maximize his principal's utility.

2.2. Private firm. Since cost savings are unobservable and the owner of the private firm is the residual claimant to the firm's profit, we have that $\lambda = 1$. The owner's maximization problem is

$$\max_{t_q, t_c} \quad \alpha t_q + \beta - C(t) + S(t_c) - \tfrac{1}{2} r \sigma_q^2 \alpha^2. \tag{2.3}$$

Let the superscript o denote the owner of the private firm. The solution to this problem is

$$\left(t_q^o, t_c^o\right) = \begin{cases} t_q = 0 \text{ and } t_c = \underline{t} \text{ s.t. } S'(\underline{t}) = C'(\underline{t}) & \text{if } 0 \leq \alpha \leq S'(\underline{t}); \\ \{(t_q, t_c) | \, \alpha = S'(t_c) = C'(t)\} & \text{if } S'(\underline{t}) < \alpha < S'(0); \\ t_q \text{ s.t. } \alpha = C'(t) \text{ and } t_c = 0 & \text{if } \alpha \geq S'(0). \end{cases}$$

In the interior solution, I get t^o, t_q^o and t_c^o as continuous functions of α with derivatives[7]

$$\frac{\partial t^o}{\partial \alpha} = \frac{1}{C''(t)} > 0; \frac{\partial t_c^o}{\partial \alpha} = \frac{1}{S''(t_c)} < 0 \text{ and } \frac{\partial t_q^o}{\partial \alpha} = \frac{1}{C''(t)} - \frac{1}{S''(t_c)} > 0.$$

In the first corner solution, t, t_q and t_c are unaffected by α. Accordingly, for $0 \leq \alpha \leq S'(\underline{t})$ we have that $t_q^o(\alpha) = 0$ and $t^o(\alpha) = t_c^o(\alpha) = \underline{t}$. Hence, if incentives for quality provision are weak, the agent only puts effort into reducing costs and total work effort is low, though still higher than under public provision. As the incentive for quality provision gets stronger, the owner increases total effort and reduces effort on cost reductions, implying that effort on quality increases. If incentives for quality provision are very strong, the owner of the private firm only puts effort into increasing quality. In addition to its effect on effort choice, α also affects the agent's risk-premium, given that there is uncertainty in the signal q (i.e. $\sigma_q^2 > 0$). The public agency's maximization problem is

$$\begin{array}{ll} \max\limits_{\alpha, \beta} & B\left(t_q^o(\alpha), \psi\right) - \left(\alpha t_q^o(\alpha) + \beta\right) \\ \text{s.t.} & \alpha \leq \overline{\alpha}, \\ & C\overline{E} \geq 0. \end{array}$$

Since the fixed wage component β is set to satisfy the owner's participation constraint, the public agency's maximization problem can be reformulated as[8]

$$\begin{array}{ll} \max\limits_{\alpha} & B\left(t_q^o(\alpha), \psi\right) - C\left(t^o(\alpha)\right) + S\left(t_c^o(\alpha)\right) - \tfrac{1}{2} r \sigma_q^2 \alpha^2 \\ \text{s.t.} & \alpha \leq \overline{\alpha}. \end{array} \tag{2.4}$$

In the interior solution, we get the first-order condition

$$\alpha^* = \frac{B_{t_q}(t_q, \psi)\left[1/C''(t) - 1/S''(t_c)\right]}{\left[1/C''(t) - 1/S''(t_c) + r\sigma_q^2\right]},$$

---

[7] It is shown in the Appendix that t^o, t_q^o and t_c^o are continuous functions of α.

[8] I show this in the Appendix.

which can be rewritten as

$$\alpha^* = \frac{B_{t_q}(t_q, \psi)}{1 + r\sigma_q^2 / \frac{\partial t_q}{\partial \alpha}}. \tag{2.5}$$

Though condition (5) seems to imply that the higher is the cost of risk, the lower is the strength of incentives for quality provision, a positive cost of risk also implies that the public agency's maximization problem need not be concave in α, and that there might be more than one interior solution.[9] Let $W(\alpha)$ denote the public agency's valuation of a certain contract, i.e. $W(\alpha) = B\left(t_q^o(\alpha)\right) - C\left(t^o(\alpha)\right) + S\left(t_c^o(\alpha)\right) - \frac{1}{2} r\sigma_q^2 \alpha^2$. Further, let $\alpha^{**} = \arg\max_{\alpha \in A} W(\alpha)$ where A is the set of feasible interior solutions.[10] To find the optimal contract, we have to compare the surplus under the corner solutions $\alpha = 0$ and $\alpha = \overline{\alpha}$ with the feasible interior solutions.[11] The optimal feasible contract is then:

$$\alpha = \begin{cases} 0 & \text{if } W(0) \geq \max\{W(\alpha^{**}), W(\overline{\alpha})\}; \\ \alpha^{**} & \text{if } W(\alpha^{**}) \geq \max\{W(0), W(\overline{\alpha})\}; \\ \overline{\alpha} & \text{if } W(\overline{\alpha}) \geq \max\{W(\alpha^{**}), W(0)\}. \end{cases}$$

3. Results

The highest possible level of quality from a public firm is simply $t_q^m = t^*$. In that case, the public agency orders the manager of the public firm to spend all his efforts on increasing quality. How high quality can be obtained from a private firm depends on the set of feasible incentives. In the one extreme, if the owner of the private firm can only be provided with weak incentives to invest in quality ($\overline{\alpha} \leq S'(\underline{t})$), it is not possible to induce the private firm to invest in quality at all. Yet if strong incentives on quality provision are feasible (that is, if $\overline{\alpha}$ is high), the maximum feasible level of quality from a private firm exceeds that of a public firm.

PROPOSITION 1. *If strong incentives are feasible, in the sense that the permissible incentive intensity $\overline{\alpha}$ exceeds some finite threshold level $\widehat{\alpha}$, a private firm can be induced to produce a higher quality than the public firm. If $\overline{\alpha}$ instead falls short of $\widehat{\alpha}$, a private firm cannot be induced to produce as high quality as a public firm.*

---

[9] Non-concavity could be ruled out by assuming that $C'''(t) \geq 0$ and $S'''(t_c) \geq 0$. This is shown in the Appendix.

[10] Note that the set of feasible interior solutions might be empty.

[11] $W(\alpha)$ is a continuous function since all the arguments are continuous in α (as shown in the proof to Proposition 1). As the derivatives of both t_q and t_c are zero for $\alpha < S'(\underline{t})$, it follows that the derivative of $W(\alpha)$ is negative for all $\alpha < S'(\underline{t})$. Hence, whenever the public agency's optimal solution implies $t_q = 0$, we have $\alpha = 0$.

Proofs are provided in the Appendix.

The result that private firms can have either higher or lower quality depending on the economic environment is also present in Hart, Shleifer and Vishny (1997), but their argument is different from mine. In their paper, the difference in quality depends on the relative importance of the quality-enhancing innovation and the negative effect on quality of cost reductions, whereas in my case it depends on the set of feasible incentives. Moreover, the difference between private and public production in Proposition 1 only refers to the *maximum* level of quality; whether a potential difference is realized hinges on the α set by the public agency. Let us now focus on differences in cost.

First, note that private provision is always cheaper than public provision for the lowest level of quality, $t_q = 0$. In that case, the manager of the public firm sets $t_c^m = t^*$ and total cost is $C(t^*) - S(t^*)$. Since $C'(t^*) = 0 < S'(t^*)$, this is clearly a suboptimally low level of effort. In contrast, the owner of the private firm is the residual claimant to cost savings and sets $t_c^o = \underline{t} > t^*$ which is the efficient level of effort. Moreover, the marginal cost for the public agency to increase quality is always lower under private provision than under public provision if the cost of risk $\left(r\sigma_q^2\right)$ is sufficiently small. To see this, note that for any given value of t_q we have that $t_c^o > t_c^m$ which by strict concavity of the cost-saving function implies that $S'(t_c^o) < S'(t_c^m)$. The cost of additional quality in terms of forgone cost savings is thus always larger for public than for private firms. However, the cost advantage of private provision vanishes for high levels of quality when the cost of risk is high. Since the public agency must increase α to induce the private firm to exert more effort on quality, the risk premium $\left(\frac{1}{2}r\sigma_q^2\alpha^2\right)$ rises with the level of quality. For high enough values of $r\sigma_q^2$, the marginal cost of quality is therefore higher under private provision than under public provision for any feasible level of quality. There is also a fixed cost of going from zero to positive levels of quality under private provision. For $0 < \alpha \leq S'(\underline{t})$, the owner of the private firm do not increase effort on t_q as α increases, but costs increase due to the higher the risk premium. Since private firms are always cheaper for zero quality, private provision is cheaper than public firms for low levels of quality, but more expensive for high levels of quality when $r\sigma_q^2$ is high.[12]

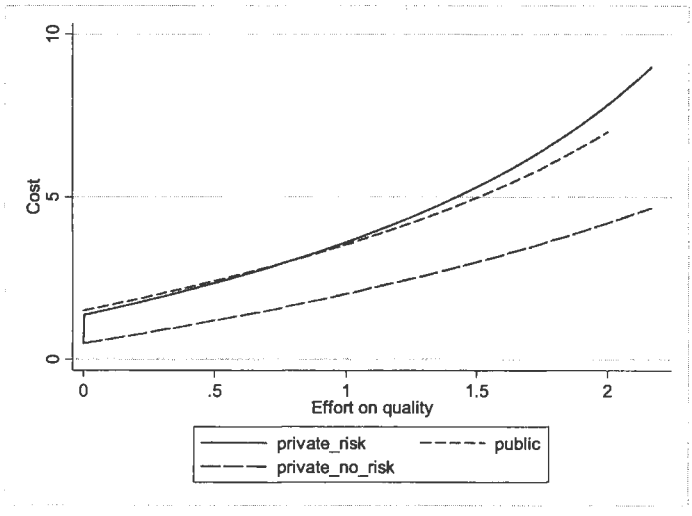
PROPOSITION 2. *There is a cost of risk $\widehat{r\sigma}_q^2$ such that private firms are cheaper for no effort on quality, but more expensive for all levels of quality above a finite threshold $\widehat{t}_q$. If $\widehat{t}_q > 0$, private firms are cheaper for all quality levels below $\widehat{t}_q$.*

Figure 1 gives two parametric examples with different levels of cost of risk. In the first case, we set $r\sigma_q^2 = 0$ and private provision is cheaper for all levels of effort on

[12] Note that, in the extreme case, "low levels of quality" refers to $t_q = 0$.

quality and the difference is increasing in the level of quality. The cost for no effort on quality is unchanged when we assume a positive cost of risk $\left(r\sigma_q^2 = 0.45\right)$, but exhibits a discontinuous jump when quality is increased above zero. The reason for this is that the strength of incentives for quality provision must be increased from $\alpha = 0$ to $\alpha = S'(\underline{t}) + \epsilon$, where ϵ is some very small number larger than zero. In this particular case, costs are similar under public and private provision for medium levels of quality, but private provision is more expensive for high levels of quality. However, the highest levels of quality can only be obtained from a private firm. If we were to increase the cost of risk even higher, we would eventually reach the case where private provision is cheaper for zero effort on quality, but more expensive for all positive levels of quality.

FIGURE 1. Parametric examples



Now, let us discuss the public agency's choice of ownership structure. Assuming that strong incentives are feasible, whether the service will be bought from a private or a public firm depends on the public agency's willingness to pay for quality, ψ, and the cost of risk, $r\sigma_q^2$. If either ψ or $r\sigma_q^2$ are low, private provision will be superior since it is cheapest. Conversely, if the public agency has a high valuation of quality and $r\sigma_q^2$ is high, a rational public agency prefers public provision since the risk premium makes private provision more expensive for high levels of quality. However, since it is not possible to extract quality in excess of t_q^* from public firms, private provision is better if the public agency's willingness to pay for quality is very high. There is thus a

potential non-monotonicity regarding the choice of ownership structure in government preference for quality.

PROPOSITION 3. *If the upper bound of contractibility on quality, $\overline{\alpha}$, is such that an effort on quality exceeding t_q^* can be extracted from private firms, private provision is superior if quality is either considered unimportant or very important, and if $r\sigma_q^2$ is sufficiently large there exists some intermediate level of the importance of quality such that public provision is superior.*

If strong incentives are not feasible, the only advantage of private provision is its lower cost for low levels of quality.

PROPOSITION 4. *If the upper bound of contractibility on quality, $\overline{\alpha}$, is such that the maximum level of quality under private provision is below t_q^*, public provision is superior when quality is important.*

To summarize, we have two cases where public ownership may be superior to private ownership, namely: i) when weak feasible incentives for effort on quality make private provision above a certain quality level impossible, and ii) when the difficulty in measuring quality makes private firms more expensive for high levels of quality.

The model illustrates the importance of controlling for market selection in empirical studies of privatization. For services where quality is hard to measure, only public agencies with a low (or very high) valuation of quality will go for private provision. Hence, unless quality is adequately controlled for, empirical studies are likely to overstate the cost-saving effect of private provision.[13] In contrast, we would expect privatization to reduce costs without any deteriation of quality for services where quality is easy to measure.[14]

Another reason why privatization might not save costs is that agency problems within private firms make the incentives for top executives similar to those of public managers. If a firm owner that hires a manager to run his business faces the same commitment problem as the government, privatization does not change anything in the context of this model.[15] Hence, privatization might have its largest impact when private

[13] Of course, the opposite case would hold if many public agency's have a very strong preference for quality and high-powered incentives for quality are feasible.

[14] There is indeed substantial evidence that privatization reduces costs without a detrimental effect on quality for services such as refuse collection and cleaning services, where measurability of quality can be assumed to be good. See, for example, Donahue (1989), Wilson (1989), Domberger and Jensen (1997) and the references citet in these works. Even a critic of privatization such as Elliot Sclar (2000) argues that "privatization of many blue-collar services certainly may be cost effective" (p. 57). However, Ohlsson (2003) argues that much of the literature on refuse collection has overstated the cost saving effect of privatization due to a failure to control for selection bias.

[15] Note that low incentives to reduce costs will imply less of an incentive balance problem, making it easier for the government to induce private firms not to shirk on quality provision.

firms are relatively small and owner-led and economic models focusing on the impact of ownership (like this paper) might only be applicable to a subset of privatization cases.

Finally, let us briefly consider the differences between contracting out in the private and public sector. One might argue that the problem of providing managers with strong incentives is worse in the public than in the private sector, since the latter can rely on the information given by stock prices. This would imply that contracting out is particularly important in the public sector. However, there are also reasons for why contracting could be more difficult in the public sector. For example, procurement laws might make implicit contracting through repeated interaction harder and contracts with private providers are often in practice enforced by government bureaucrats with weak incentives to do a good job.

4. Extensions

4.1. Contracting on cost. I have so far analyzed the case where it is only possible to contract on quality. Here I relax this assumption and let the public agency and the private firm contract on cost.[16] As we will see, because of the residual risk, it is not possible to replicate the public firm even under full contractibility on quality and cost.

Under no cost contracting, the owner of the firm is the residual claimant to cost savings and $\lambda = 1$. Hence, cost contracting implies that the firm and the government sign a contract that gives the private firm *less* than the full share of cost savings, in effect a cost-plus contract where the firm gets compensation for costs incurred.[17] Accordingly, complete cost-plus contracting corresponds to $\lambda = 0$. The cost contracting problem is thus not to provide private firms with strong incentives to cut costs, but to mitigate those incentives, and the commitment problem is to make a high degree of compensation credible. In terms of our model, the *lower* is t_c, the *larger* is the compensation from the government to the private firm.

Let $\gamma = 1 - \lambda$ denote the share of costs for which the owner is compensated. I assume that $\gamma \in [0,1]$, though in principle there might be some upper bound on contractibility on cost $\overline{\gamma} \in [0,1]$ so that $\gamma \in [0,\overline{\gamma}]$. The principal receives a signal of costs $v = S(t_c) + \varepsilon_v$ where $\varepsilon_v \sim N(0,\sigma_v^2)$ and the agent's compensation is $w(q,v) = \alpha q + \beta - \gamma v$. I assume that the errors in q and v are independent, i.e., $Cov(\varepsilon_q,\varepsilon_v) = 0$. The owner's monetary payoff, x, is then normally distributed with $E(x) = \alpha t_q + \beta -$

[16] I keep the assumption that no contract on the completion of tasks is possible under public provision.

[17] I assume that it is not possible for the owner of the private firm to extract revenue by increasing costs, for example by increasing his own wage.

$C(t)+(1-\gamma)S(t_c)$ and $V(x)=\sigma_q^2\alpha^2+\sigma_v^2\gamma^2$. This gives us the owner's maximization problem

$$\max_{t_q,t_c} \alpha t_q+\beta-C(t)+(1-\gamma)S(t_c)-\frac{1}{2}r\left(\sigma_q^2\alpha^2+\sigma_v^2\gamma^2\right). \tag{4.1}$$

The solution to this problem is

$$\left(t_q^o,t_c^o\right)=\begin{cases}\{(t_q,t_c)|\, t_q+t_c=t^*\} & \text{if } \alpha=0 \text{ and } \gamma=1;\\ t_q=0 \text{ and } t_c=\underline{t} \text{ s.t. } (1-\gamma)S'(\underline{t})=C'(\underline{t}) & \text{if } 0\le\alpha\le(1-\gamma)S'(\underline{t}) \text{ and } \gamma<1;\\ \{(t_q,t_c)|\, \alpha=(1-\gamma)S'(t_c)=C'(t)\} & \text{if } (1-\gamma)S'(\underline{t})<\alpha<(1-\gamma)S'(0);\\ \{(t_q,t_c)|\, \alpha=C'(t), t_c=0\} & \text{if } \alpha>(1-\gamma)S'(0).\end{cases}$$

In the interior solution (excluding the case when $\alpha=0$ and $\gamma=1$), we get t^o, t_q^o and t_c^o as continuous functions of α and γ with partial derivatives[18]

$$\begin{aligned}\frac{\partial t^o}{\partial\alpha} &= \frac{1}{C''(t)}>0\ ;\ \frac{\partial t^o}{\partial\gamma}=0\ ;\\ \frac{\partial t_c^o}{\partial\alpha} &= \frac{1}{(1-\gamma)S''(t_c)}<0\ ;\ \frac{\partial t_c^o}{\partial\gamma}=\frac{[S'(t_c)]^2}{\alpha S''(t_c)}<0\ ;\\ \frac{\partial t_q^o}{\partial\alpha} &= \frac{1}{C''(t)}-\frac{1}{(1-\gamma)S''(t_c)}>0 \text{ and } \frac{\partial t_q^o}{\partial\gamma}=-\frac{[S'(t_c)]^2}{\alpha S''(t_c)}>0.\end{aligned}$$

The public agency's maximization problem is therefore

$$\begin{array}{ll}\max\limits_{\alpha,\gamma} & B\left(t_q^o(\alpha,\gamma)\right)-C\left(t^o(\alpha)\right)+S\left(t_c^o(\alpha,\gamma)\right)-\frac{1}{2}r\left(\sigma_q^2\alpha^2+\sigma_v^2\gamma^2\right)\\ \text{s.t.} & \alpha\le\overline{\alpha},\\ & \gamma\in[0,1].\end{array}$$

We get the first-order condition with respect to α,

$$\alpha^*=\frac{B_{t_q}(t_q,\psi)}{1+r\sigma_q^2/\frac{\partial t_q^o}{\partial\alpha}}, \tag{4.2}$$

and the first-order condition with respect to γ,

$$\gamma^*=\frac{B_{t_q}(t_q,\psi)-S'(t_c)}{r\sigma_v^2}\frac{\partial t_q^o}{\partial\gamma}. \tag{4.3}$$

As above, the maximization problem is not necessarily concave in α and γ.

The basic insight from this extension of the model is that the level of incentives for quality provision (α) needed to ensure a certain level of quality (t_q) is falling in the extent of cost-plus contracting (γ). By compensating the owner of the firm for his costs, incentives to invest in cost reductions are muted, implying that effort on the cost reducing task falls, which makes the cost of effort for engaging in quality provision lower

[18] This is shown in the Appendix.

on the margin.[19] Hence, by lowering the incentives for cost reductions, a higher level of quality can be exerted for any given level of incentives for quality provision. This has two potential advantages: First, it mitigates the problem of weak feasible incentives for quality provision. Second, for a given level of quality, incentives for quality can be less high-powered, thereby lowering the cost of risk.[20]

Note that setting $\alpha = 0$ and $\gamma = 1$ gives the owner of the private firm exactly the same incentives as the manager of the public firm. However, if $r\sigma_v^2 > 0$, the owner of the private firm have to bear risk, making private provision more expensive. Hence, ownership matters also when a private firm can be provided with exactly the same incentives as a public firm. Actually, even if contracts are complete (in the sense that $\overline{\alpha} = \infty$ and $\overline{\gamma} = 1$), public provision is still cheaper than private provision for high levels of quality if $r\sigma_q^2$ and $r\sigma_v^2$ are high enough.

PROPOSITION 5. *For every positive level of quality there exists a nonempty set of vectors* $\left(\widehat{r\sigma}_q^2, \widehat{r\sigma}_v^2\right)$ *such that public firms are cheaper than private firms.*

4.2. Task attractiveness. A key result of the model is that private firms underinvest in tasks for which they have no financial incentives, whereas the drawback of public provision is the low level of total effort. This is in accordance with previous models of privatization like Schmidt (1996) and Hart, Shleifer and Vishny (1997). An implicit assumption in these models, as well as my model, is that tasks do not differ in terms of their inherent attractiveness. Yet if some tasks are inherently less attractive than others (for example because they are boring or difficult to undertake), the incentive balance problem might be reversed. The manager of a public firm, who lacks financial incentives for any particular task, underinvests in those tasks that he or she for some reason finds unattractive. The owner of the private firm also dislikes the unattractive task, but this is at least partly offset by an explicit financial incentive.

Let t_u denote effort on the unattractive task. I assume that t_u, t_q and t_c are perfect substitutes in the cost of effort, i.e., $t = t_u + t_q + t_c$. Effort on t_u also comes at an additional cost of effort $D(t_u)$, which is an increasing and strictly convex function with $D'(0) > 0$. I assume that effort on t_u is perfectly observable and hence contractible in case of private provision. The owner of the private firm is compensated according to the linear wage contract

[19] As long as $\alpha > (1-\gamma)S'(\underline{t})$, total effort is determined by α, and γ does only determine the allocation across tasks – the higher is γ, the larger must $S'(t_c)$ be for a given α, implying that t_c is falling in γ. For $0 \leq \alpha \leq (1-\gamma)S'(\underline{t})$, total effort is determined solely by γ, but then there's no rationale for setting $\gamma > 0$, except for eliminating all incentives for cost reductions by setting $\gamma = 1$

[20] Conditions under which the principal will set $\gamma > 0$ are provided in the Appendix.

$$w(q) = \alpha q + \omega t_u + \beta.$$

The owner's maximization problem then becomes

$$\max_{t_q,\, t_c,\, t_u} \quad \alpha t_q + \omega t_u + \beta + S(t_c) - C(t) - D(t_u) - \tfrac{1}{2} r \sigma_q^2 \alpha^2. \tag{4.4}$$

Solving this problem, we find that the owner of the private firm will put effort on the unattractive task if $\omega > D'(0) + \max\{\alpha, S'(\underline{t})\}$. Note that effort on cost reductions and quality are both crowded out by strong incentives for the unattractive task.

Since the public agency cannot commit to any incentive scheme under public provision, we have $\alpha = 0$ and $\omega = 0$, implying that the public manager's maximization problem is

$$\max_{t_q, t_c, t_u} \quad \beta - C(t) - D(t_u). \tag{4.5}$$

Hence, under public provision, the manager sets $t_u = 0$ and $t_q + t_c = t^*$. Note that I implicitly assume that the government cannot restrict the work description of the public firm to only include t_u, in which case the task allocation problem might be overcome.[21] By the same argument, differences in task attractiveness also provides a rationale for limiting worker freedom within organizations.

PROPOSITION 6. *Public firms shun unattractive tasks, but private firms undertake them if incentives are strong enough.*

Instead of an additional unattractive task, we could consider the relative task attractiveness of t_q and t_c. For example, it is conceivable that improving quality is a more attractive task than reducing costs. If so, it would not be possible to induce a manager of a public firm to put effort on cost reductions. This is an important qualification to the assumption above that the manager of the public firm is indifferent between improving quality and reducing costs.

5. Concluding remarks

This paper has argued that differences in ex ante contracts is an important feature of privatization of service provision. Under private provision, procurement agencies write elaborate contracts with private firms, whereas managers under in-house provision seldom have explicit incentive contracts. This is a stark view and there are of course examples of less clear-cut cases. Private firms might in effect be run by employed

[21] For the manager of the public firm to set $t_u > 0$, we must have that $|C'(0)| > D'(0)$. Note that if t_u is not an argument in $C(t)$ (that is, the cost of effort on t_q and t_c are independent of t_u), then the manager of the public firm will always set $t_u = 0$.

managers with only a limited stake in the firm's profits. Public agencies might create independent firms that have to compete with private service providers for contracts. Such arrangements could reduce, or even eliminate, differences between private and public provision. However, it might still be fruitful to analyze such intermediate cases as examples of different ex ante contracts.

Appendix A

A.1. Owner's effort choice without contracting on costs. Consider the owner's effort on cost reductions as a function of α. From the solution to (3), we know that $t_c^o(\alpha) = \underline{t} > t^*$ for $0 \leq \alpha \leq S'(\underline{t})$ and $t_c^o(\alpha) = 0$ for all $\alpha \geq S'(0)$. Since $S(t_c)$ is increasing and strictly concave, there exists an inverse function to $\alpha = S'(t_c)$ for $0 < t_c < \underline{t}$. As $S(t_c)$ is also twice continuously differentiable, the inverse function $t_c^o(\alpha)$ is continuously decreasing in α for $S'(\underline{t}) < \alpha < S'(0)$. Now consider the owner's total effort as a function of α. From the solution to (3), we know that $t_q^o(\alpha) = 0$ for $0 \leq \alpha \leq S'(\underline{t})$ and so $t^o(\alpha) = t_c^o(\alpha) = \underline{t}$ for $0 \leq \alpha \leq S'(\underline{t})$. Since $C(t)$ is strictly convex and increasing for $t > t^*$, $\alpha = C'(t)$ is invertible for $\underline{t} < t < \bar{t}$ where $\bar{t}$ is such that $\lim_{t \to \bar{t}} C'(t) = \infty$. As $C(t)$ is also twice continuously differentiable, the inverse function $t^o(\alpha)$ is continuously increasing in α for $\alpha > S'(\underline{t}) = C'(\underline{t})$ with $\lim_{\alpha \to \infty} t^o(\alpha) = \bar{t}$. The owner's effort on quality is given by $t_q^o(\alpha) = t^o(\alpha) - t_c^o(\alpha)$. It follows from above that $t_q^o(\alpha) = 0$ for $0 \leq \alpha \leq S'(\underline{t})$ and that $t_q^o(\alpha)$ is continuously increasing in α for all $\alpha > S'(\underline{t}) = C'(\underline{t})$ with $\lim_{\alpha \to \infty} t_q^o(\alpha) = \bar{t}$.

A.2. Owner's effort choice with contracting on costs. For $\alpha = 0$ and $\gamma = 1$, the owner of the firm is indifferent between t_q and t_c and all allocations of effort such that $t_q + t_c = t^*$ are feasible. However, as soon as $\alpha > 0$ and $\gamma = 1$, the owner will set $t_c = 0$ or, alternatively, when $\alpha = 0$ and $\gamma < 1$, the owner will set $t_q = 0$. Hence, $t_q^o(\alpha, \gamma)$ and $t_c^o(\alpha, \gamma)$ are not necessarily continuous in α and γ when $\alpha = 0$ and $\gamma = 1$.

From the solution to (6), it follows that for $\gamma < 1$, $t^o(\alpha, \gamma)$, $t_c^o(\alpha, \gamma)$ and $t_q^o(\alpha, \gamma)$ are continuous in α by the same argument as in the case without contracting on costs (to see this, just replace α by $\alpha/(1-\gamma)$ in the optimality conditions). For $\gamma = 1$, $t^o(\alpha, \gamma)$ and $t_q^o(\alpha, \gamma)$ are continuously increasing in α.

From the solution to (6), we see that $t_q = 0$ for $0 \leq \frac{\alpha}{1-\gamma} \leq S'(\underline{t})$, $t_c^o(\alpha, \gamma) = 0$ for $\frac{\alpha}{1-\gamma} \geq S'(0)$ and that $t^o(\alpha, \gamma)$ is independent of γ for $\frac{\alpha}{1-\gamma} > S'(\underline{t})$. Since $S(t_c)$ is increasing and strictly concave, there exists an inverse function to $\frac{\alpha}{1-\gamma} = S'(t_c)$ for $0 < t_c < \underline{t}$. As $S(t_c)$ is also twice continuously differentiable, the inverse function $t_c^o(\alpha, \gamma)$ is continuously decreasing in $\frac{\alpha}{1-\gamma}$ for $S'(\underline{t}) < \frac{\alpha}{1-\gamma} < S'(0)$ and hence continuously decreasing in γ for fixed α. It follows that $t_q^o(\alpha, \gamma) = t^o(\alpha, \gamma) - t_c^o(\alpha, \gamma)$ is continuously increasing in γ for $S'(\underline{t}) < \frac{\alpha}{1-\gamma} < S'(0)$ and independent of γ for $0 \leq \frac{\alpha}{1-\gamma} \leq S'(\underline{t})$ and $\frac{\alpha}{1-\gamma} \geq S'(0)$.

A.3. The public agency's maximization problem. The public agency's maximization problem without contracting on costs is

$$\begin{array}{ll}\max\limits_{\alpha,\beta} & E\left[B\left(t_q^o(\alpha),\psi\right)-w(q)\right]\\ \text{s.t.} & \alpha\leq\overline{\alpha},\\ & CE\geq 0.\end{array}$$

Since the public agency is risk neutral, we get

$$\begin{array}{ll}\max\limits_{\alpha,\beta} & B\left(t_q^o(\alpha),\psi\right)-(\alpha t_q+\beta)\\ \text{s.t.} & \alpha\leq\overline{\alpha},\\ & CE\geq 0.\end{array}\tag{A.1}$$

The fixed wage β is set just to let the participation constraint bind, i.e., so that $CE=0$. Reformulating the CE, we get

$$\beta=C\left(t^o(\alpha)\right)+\frac{1}{2}r\sigma_q^2\alpha^2-S\left(t_c^o(\alpha)\right)-\alpha t_q.$$

Substituting β into (11), we get

$$\begin{array}{ll}\max\limits_{\alpha} & B\left(t_q^o(\alpha),\psi\right)-\left(\alpha t_q+C\left(t^o(\alpha)\right)+\frac{1}{2}r\sigma_\varepsilon^2\alpha^2-S\left(t_c^o(\alpha)\right)-\alpha t_q\right)\\ \text{s.t.} & \alpha\leq\overline{\alpha}.\end{array}$$

This maximimization problem can be reformulated as

$$\begin{array}{ll}\max\limits_{\alpha} & B\left(t_q^o(\alpha),\psi\right)-C\left(t^o(\alpha)\right)+S\left(t_c^o(\alpha)\right)-\frac{1}{2}r\sigma_q^2\alpha^2\\ \text{s.t.} & \alpha\leq\overline{\alpha}.\end{array}$$

A.4. First and second derivatives.

A.4.1. *Without contracting on cost.* The first derivative of the public agency's maximization problem with respect to α is

$$\begin{aligned}\frac{\partial W(\alpha)}{\partial\alpha} &= B_{t_q}\left(t_q^o(\alpha),\psi\right)\left[\frac{1}{C''\left(t(\alpha)\right)}\right]-B_{t_q}\left(t_q^o(\alpha),\psi\right)\left[\frac{1}{S''\left(t_c(\alpha)\right)}\right]\\ &\quad -C'\left(t(\alpha)\right)\left[\frac{1}{C''\left(t(\alpha)\right)}\right]+S\left(t_c(\alpha)\right)\left[\frac{1}{S''\left(t_c(\alpha)\right)}\right]-r\sigma_q^2\alpha.\end{aligned}$$

The second derivative with respect to α becomes

$$\begin{aligned}\frac{\partial^2 W(\alpha)}{\partial\alpha^2} &= \underbrace{B_{t_q,t_q}(t_q,\psi)\left[\left[1/C''(t)-1/S''(t_c)\right]^2\right]}_{-}+\underbrace{\left[B_{t_q}(t_q,\psi)-S'(t_c)\right]\left[S''(t_c)\right]^{-3}S'''(t_c)}_{\pm}\\ &+\underbrace{\left[C'(t)-B_{t_q}(t_q,\psi)\right]\left[C''(t)\right]^{-3}C'''(t)}_{\pm}+\underbrace{\left[1/S''(t_c)\right]-\left[1/C''(t)\right]}_{-}-\underbrace{r\sigma_q^2}_{+}\end{aligned}$$

which can be either larger or smaller than zero. The second derivative is negative for certain if $B_{t_q}(t_q,\psi)=S'(t_c)=C'(t)$ (which is the case if $r\sigma_q^2=0$) or if $C'''(t)\geq 0$ and $S'''(t_c)\geq 0$.

A.4.2. *With contracting on cost.* The first derivative of the public agency's maximization problem with respect to α is

$$\begin{aligned}\frac{\partial W(\alpha,\gamma)}{\partial\alpha} &= B_{t_q}\left(t_q^o(\alpha,\gamma)\right)\left[\frac{1}{C''(t)}\right] - B_{t_q}\left(t_q^o(\alpha,\gamma)\right)\left[\frac{1}{(1-\gamma)S''(t_c)}\right] \\ &\quad -C'\left(t^o(\alpha)\right)\left[\frac{1}{C''(t)}\right] + S'\left(t_c^o(\alpha,\gamma)\right)\left[\frac{1}{(1-\gamma)S''(t_c)}\right] - r\sigma_q^2\alpha\end{aligned}$$

which is not defined for $\gamma = 1$. The second derivative with respect to α is

$$\begin{aligned}\frac{\partial^2 W(\alpha,\gamma)}{\partial\alpha^2} &= \underbrace{\left[B_{t_q}(t_q,\psi) - S'(t_c)\right]\left[(1-\gamma)S''(t_c(\alpha))\right]^{-3}(1-\gamma)S'''(t_c)}_{\pm} \\ &+\underbrace{\left[1/\left[(1-\gamma)^2 S''(t_c(\alpha))\right]\right] - \left[1/C''(t(\alpha))\right]}_{-} \\ &\underbrace{B_{t_q,t_q}(t_q,\psi)\left[\left[1/C''(t(\alpha)) - 1/\left[(1-\gamma)S''(t_c(\alpha))\right]\right]^2\right]}_{-} \\ &+\underbrace{\left[C'(t) - B_{t_q}(t_q,\psi)\right]\left[C''(t(\alpha))\right]^{-3}C'''(t)}_{\pm} - \underbrace{r\sigma_q^2}_{+}\end{aligned}$$

which is negative for certain if $B_{t_q}(t_q,\psi) = S'(t_c) = C'(t)$ or if $C'''(t) \geq 0$ and $S'''(t_c) \geq 0$.

The first derivative of the public agency's maximization problem with respect to γ is

$$\frac{\partial W(\alpha,\gamma)}{\partial\gamma} = -B_{t_q}(t_q,\psi)\left[\frac{[S'(t_c)]^2}{\alpha S''(t_c)}\right] + S'(t_c)\left[\frac{[S'(t_c)]^2}{\alpha S''(t_c)}\right] - \gamma r\sigma_v^2.$$

This derivative is positive for $\gamma = 0$ whenever $B_{t_q}(t_q,\psi) > S'(t_c)$.

For the public agency to set $\gamma > 0$ it must either be the case that the cost of risk is positive or that strong incentives for quality are infeasible. To see this, suppose that we restict γ to zero. The public agency's problem is then exactly the same as without contracting on cost. Further, suppose that quality is so important that the solution to the public agency's problem is such that $t_q > 0$ and that the incentive feasibility constraint $\overline{\alpha}$ does not bind. Then if $r\sigma_q^2 > 0$, we see from (7) that $\alpha = S'(t_c) < B_{t_q}(t_q,\psi)$ and so $\gamma = 0$ cannot be optimal. Similarly, γ will be strictly above zero if strong incentives for quality provision are not feasible, i.e., if $B_{t_q}\left(t_q(\overline{\alpha},0),\psi\right) > S'(t_c(\overline{\alpha},0))$, but quality is so important that $t_q(\overline{\alpha},0)$ is preferred over $t_q = 0$.

Before we consider the second derivative of the public agency's maximization problem with respect to γ, note that the derivative of

$$\left[\frac{[S'(t_c)]^2}{\alpha S''(t_c)}\right]$$

with respect to γ can be written as

$$\begin{aligned}
&\frac{1}{\alpha}\left[2\left[S'\left(t_c\right)\right]\left[S''\left(t_c\right)\right]^{-1}S''\left(t_c\right)\frac{\left[S'\left(t_c\right)\right]^2}{\alpha S''\left(t_c\right)}+\left[S'\left(t_c\right)\right]^2(-1)\left[S''\left(t_c\right)\right]^{-2}S'''\left(t_c\right)\frac{\left[S'\left(t_c\right)\right]^2}{\alpha S''\left(t_c\right)}\right] \\
= &\frac{1}{\alpha^2}\left[2\left[S'\left(t_c\right)\right]^3\left[S''\left(t_c\right)\right]^{-2}S''\left(t_c\right)-\left[S'\left(t_c\right)\right]^4\left[S''\left(t_c\right)\right]^{-3}S'''\left(t_c\right)\right] \\
= &\frac{\left[S'\left(t_c\right)\right]^3}{\alpha^2 S''\left(t_c\right)}\left[2-\frac{S'\left(t_c\right)S'''\left(t_c\right)}{\left[S''\left(t_c\right)\right]^2}\right]\gtrless 0.
\end{aligned}$$

Using this, we can write the second derivative of the owner's maximization problem with respect to γ as

$$\begin{aligned}
\frac{\partial^2 W(\alpha,\gamma)}{\partial\gamma^2} = &\underbrace{\left[S'\left(t_c\right)-B_{t_q}\left(t_q,\psi\right)\right]\left[\frac{\left[S'\left(t_c\right)\right]^3}{\alpha^2 S''\left(t_c\right)}\left[2-\frac{S'\left(t_c\right)S'''\left(t_c\right)}{\left[S''\left(t_c\right)\right]^2}\right]\right]}_{\pm} \\
&\underbrace{+\left[S''\left(t_c\right)+B_{t_q,t_q}\left(t_q,\psi\right)\right]\left[\frac{\left[S'\left(t_c\right)\right]^2}{\alpha S''\left(t_c\right)}\right]^2}_{-}-\underbrace{r\sigma_v^2}_{+}.
\end{aligned}$$

Note that it can never be optimal to set (α,γ) such that $S'\left(t_c\right)>B_{t_q}\left(t_q,\psi\right)$, since the public agency then can achieve both a better allocation of the agents effort and reduce the cost of risk by reducing α or γ. Since $B_{t_q}\left(t_q,\psi\right)\geq S'\left(t_c\right)$, a sufficient condition for the second derivative to be negative is that

$$\frac{S'\left(t_c\right)S'''\left(t_c\right)}{\left[S''\left(t_c\right)\right]^2}\geq 2.$$

Finally, the cross-partial derivative is

$$\begin{aligned}
\frac{\partial^2 W(\alpha,\gamma)}{\partial\alpha\partial\gamma} = \frac{\partial^2 W(\alpha,\gamma)}{\partial\gamma\partial\alpha} = &\underbrace{\left[B_{t_q}\left(t_q\right)-S'\left(t_c\right)\right]\left[\frac{\left[S'\left(t_c\right)\right]^2\frac{S'''\left(t_c\right)}{S''\left(t_c\right)}-S'\left(t_c\right)S''\left(t_c\right)}{\alpha\left(1-\gamma\right)\left[S''\left(t_c\right)\right]^2}\right]}_{\pm} \\
&\underbrace{+\left[\frac{B_{t_q t_q}\left(t_q\right)+S''\left(t_c\right)}{S''\left(t_c\right)}\right]\frac{\left[S'\left(t_c\right)\right]^2}{\alpha\left(1-\gamma\right)S''\left(t_c\right)}}_{-}-\underbrace{\frac{B_{t_q t_q}\left(t_q\right)}{C''\left(t\right)}\frac{\left[S'\left(t_c\right)\right]^2}{\alpha S''\left(t_c\right)}}_{+}.
\end{aligned}$$

Since the cross-partial is not zero, it is a nontrivial exercise to derive conditions under which the Hessian matrix is negative definite. As it is not important for any of the results presented in the paper, I abstain from deriving these conditions here.

A.5. Proof of Proposition 1. From (1), we know that the maximum level of quality under public provision is $t_q^m = t^*$. Since $\bar{t} > t^*$, there exists some $S'(\underline{t}) < \widehat{\alpha} < \infty$ by continuity of $t_q^o(\alpha)$ such that $t_q^o(\widehat{\alpha}) = t^*$, $t_q^o(\alpha) < t^*$ for all $\alpha < \widehat{\alpha}$ and $t_q^o(\alpha) > t^*$ for all $\alpha > \widehat{\alpha}$.

A.6. Proof of Proposition 2. Consider the function $t_q^o(\alpha)$ from the owner's effort choice above. As $t_q^o(\alpha)$ is continuously increasing in α for all $\alpha > S'(\underline{t})$, it has an inverse function $\alpha\left(t_q^o\right)$ for $t_q^o > 0$ which is continuously increasing in t_q^o. From the solution to (4) we also know that $\alpha(0) = 0$. Total costs for a given level of quality under private provision is then given by

$$K_o(\alpha(t_q)) = C(t^o(\alpha(t_q))) - S(t_c^o(\alpha(t_q))) + \frac{1}{2}\left[r\sigma_q^2(\alpha(t_q))^2\right]$$

for $0 \leq t_q \leq \bar{t}$. As $\alpha\left(t_q^o\right)$, $C(t)$ and $S(t_c)$ are continuous functions for $t_q > 0$, $K_o(\alpha(t_q))$ is a continuous function of t_q for $t_q > 0$.

Total costs under public provision is given by

$$K_m(t_q) = C(t^*) - S(t^* - t_q)$$

for $0 \leq t_q \leq t_q^*$, which is a continuous function by continuity of $S(t_c)$. First, consider the cost for no effort on quality. In case of private provision, we get

$$\begin{aligned} K_o(0) &= C(t^o(0)) - S(t_c^o(0)) \\ &= C(\underline{t}) - S(\underline{t}) \end{aligned}$$

which is the amount of effort that minimizes K. Since t^* is such that $C(t^*) = 0 < S(t^*)$, it follows that

$$K_o(0) < K_m(0).$$

Second, consider the cost for $t_q > 0$. In case of private provision, we get

$$K_o\left(\alpha\left(t_q^o\right)\right) = C\left(t^o\left(\alpha\left(t_q^o\right)\right)\right) - S\left(t_c^o\left(\alpha\left(t_q^o\right)\right)\right) + \frac{1}{2}\left[r\sigma_q^2(\alpha(t_q))^2\right].$$

Since $C(t^o(\alpha(t_q))) - S(t_c^o(\alpha(t_q)))$ is a finite number and $\alpha\left(t_q^o\right) > 0$ for $t_q^o > 0$, it follows that $\lim_{r\sigma_q^2 \to \infty} K_o\left(\alpha\left(t_q^o\right)\right) = \infty$ for all $t_q^o > 0$. In contrast, total costs under public provision is given by

$$K_m(t_q) = C(t^*) - S(t^* - t_q)$$

which is a finite number for all $t_q \leq t_q^*$. By continuity of $K_o(\alpha(t_q))$ and $K_m(t_q)$ for $t_q > 0$, there exists a finite value $\widehat{r\sigma}_q^2$ for all $0 < t_q \leq t_q^*$ such that, $K(t_q)_o > K(t_q)_m$ whenever $r\sigma_q^2 > \widehat{r\sigma}_q^2$.

Third, consider the marginal cost of quality under private provision for $t_q > 0$,

$$\frac{\partial K(\alpha(t_q))}{\partial t_q} = \frac{\partial K}{\partial \alpha}\frac{\partial \alpha}{\partial t_q} = \left(\frac{C'(t^o(\alpha(t_q)))}{C''(t(\alpha(t_q)))} - \frac{S'(t_c^o(\alpha(t_q)))}{S''(t_c(\alpha(t_q)))} + \alpha r\sigma_\varepsilon^2\right)\frac{\partial \alpha}{\partial t_q}.$$

Since

$$\frac{C'(t^o(\alpha(t_q)))}{C''(t(\alpha(t_q)))} - \frac{S'(t_c^o(\alpha(t_q)))}{S''(t_c(\alpha(t_q)))}$$

is positive, and

$$\frac{\partial \alpha}{\partial t_q} > 0,$$

it follows that

$$\lim_{r\sigma_q^2 \to \infty} \frac{\partial K_o(\alpha(t_q))}{\partial t_q} = \infty.$$

Since

$$\frac{\partial K_m(t_q)}{\partial t_q} = S'(t^* - t_q)$$

is finite for all $t_q \leq t^*$, for $t_q > 0$ there exists a finite value $\widetilde{r\sigma}_q^2$ such that

$$\frac{\partial K_o(\alpha(t_q))}{\partial t_q} > \frac{\partial K_m(t_q)}{\partial t_q}$$

if $r\sigma_q^2 > \widetilde{r\sigma}_q^2$.

It follows that there exists some $r\sigma_q^2 > \max\{\widehat{r\sigma}_q^2, \widetilde{r\sigma}_q^2\}$ such that private provision is cheaper for zero effort on quality but more expensive for effort on quality above some finite threshold $t_q < t_q^*$. As marginal cost is higher under private provision for $t_q > 0$, costs can be the same for at most one level of effort on quality.

A.7. Proof of Proposition 3. I first show that public provision can be superior. As $K_m(t_q^m) - K_o(0)$ is a finite number and $B(t_q, \psi)$ is continuous in ψ with $\lim_{\psi \to \infty} B(t_q, \psi) = \infty$, there always exists some finite $\widetilde{\psi}$ such that $B(t_q^m, \psi) - K_m(t_q^m) > B(0, \psi) - K_o(0)$ for all $\psi > \widetilde{\psi}$. That is, if quality is sufficiently important, the optimal level of quality from public provision is superior to zero quality from a private firm. We know from Proposition 2 that $\lim_{r\sigma_q^2 \to \infty} K_o(t_q) = \infty$ for all $t_q^o > 0$. It follows that the surplus under private provision approaches $B(0, \psi) - K_o(0)$ from above when $r\sigma_q^2$ goes to infinity. Hence, for all $\psi > \widetilde{\psi}$, there also exists a finite cost of risk $\widehat{r\sigma}_q^2(\psi)$ such that public provision is preferred to private provision for $r\sigma_q^2 > \widehat{r\sigma}_q^2(\psi)$.

Second, consider the case when quality is unimportant ($\psi = 0$). Since $B_{t_q}(t_q, 0) = 0$ for all t_q we get from the solutions to (2) and (4) that the optimal level of t_q is equal to zero both under private and public provision. From Proposition 2, we know that $K_o(0) < K_m(0)$, and so private provision is superior.

Before we consider the case when quality is very important, we establish that there exists some finite ψ such that the public agency will set α so that $t_q^o(\alpha) > t^* \geq t_q^m$. As $\lim_{\psi \to \infty} B(t_q, \psi) = \infty$ and $K_o(t_q^o)$ is a finite number for any $t_q < \bar{t}$, there exists for all $0 < t_q < \bar{t}$ some finite value of ψ such that $B(t_q, \psi) - K_o(t_q) > B(0, \psi) - K_o(0)$. We can therefore rule out the corner solution $\alpha = 0$. Since $\lim_{\psi \to \infty} B_{t_q}(t_q, \psi) = \infty$, we get from (5) that $\lim_{\psi \to \infty} \alpha^* = \infty$. As $\lim_{\alpha \to \infty} t_q^o(\alpha) = \bar{t}$, we have that $\lim_{\psi \to \infty} t_q^o = \bar{t}$. To see that private provision is superior when quality is very important, note that since $K_o(t_q^o) - K_m(t_q^m)$ is a finite number, $\lim_{\psi \to \infty} B_{t_q}(t_q, \psi) = \infty$ for all $t_q > 0$ and $\lim_{\psi \to \infty} t_q^o = \bar{t} > t^* \geq t_q^m$, there always exists a $\widehat{\psi}(r\sigma_q^2)$ such that $B(t_q^o, \psi) - K_o(t_q^o) > B(t_q^m, \psi) - K_m(t_q^m)$ for all $\psi > \widehat{\psi}$.

Now, suppose that $r\sigma_q^2 > \widehat{r\sigma}_q^2(\psi)$ for some $\psi > \widetilde{\psi}$. We know from above that private provision is superior if $\psi = 0$ or $\psi > \widehat{\psi}(r\sigma_q^2)$ and that there exists some ψ above $\widetilde{\psi}$ but below $\widehat{\psi}(r\sigma_q^2)$ such that public provision is superior.

A.8. Proof of Proposition 4. We know from Proposition 1 that there exists an $\widehat{\alpha}$ such that $t_q^o(\overline{\alpha}) < t_q^*$ if $\overline{\alpha} < \widehat{\alpha}$. Since $\lim_{\psi \to \infty} B_{t_q}(t_q, \psi) = \infty$ and $t_q^m = t^*$ if $B_{t_q}(t_q, \psi) \geq S'(0)$, there exists a finite $\widetilde{\psi}$ such that $t_q^m > t_q^o$ if $\psi > \widetilde{\psi}$ and $\overline{\alpha} < \widehat{\alpha}$. Since $K_o(t_q^o) - K_m(t_q^m)$ is finite and $\lim_{\psi \to \infty} B_{t_q}(t_q, \psi) = \infty$ there exists a $\widehat{\psi}$ such that $B(t_q^m, \psi) - B(t_q^o, \psi) > K_m(\overline{t_q}) - K_o(\overline{t_q})$ for all $\psi > \widehat{\psi}$.

A.9. Proof of Proposition 5. For any feasible level of effort on quality, there is some nonempty set of parameter values (α, γ) that induces the private firm to put down this particular level of effort (the particular parameter values that minimize costs for a certain level of quality will vary depending on $r\sigma_q^2$ and $r\sigma_v^2$). Let $\alpha(t_q)$ and $\gamma(t_q)$ denote the values of the incentive parameters that minimize total cost to the public agency for a certain level of t_q. Then, total cost for a given level of quality is

$$K(t_q) = C(t^o(\alpha(t_q), \gamma(t_q))) - S(t_c^o(\alpha(t_q), \gamma(t_q))) + \frac{1}{2} r\left(\sigma_q^2(\alpha(t_q))^2 + \sigma_v^2(\gamma(t_q))^2\right).$$

Note that $(\alpha, \gamma) = (0, 0)$ is the only optimal solution for $t_q = 0$. We therefore get the same result as in Proposition 2 regarding cost for zero effort on quality, i.e., $K_o(0, 0) = K(0)_o < K(0)_m$.

Second, for any $t_q > 0$, we must have that either α and γ or both are strictly larger than zero. Since $C(t^o(\alpha)) - S(t_c^o(\alpha,\gamma))$ is finite, costs approach infinity for the whole set of parameter values that give a certain t_q (including the set that minimize costs) when both $r\sigma_q^2$ and $r\sigma_v^2$ approach infinity. Hence, we have that $\lim_{r\sigma_q^2\to\infty; r\sigma_v^2\to\infty} K_o(t_q) = \infty$ for all feasible $t_q > 0$. Since costs are continuously increasing in $r\sigma_q^2$ and $r\sigma_v^2$ for all combinations of α and γ that give a certain level of quality, $K_o(t_q)$ is a continuous function of $r\sigma_q^2$ and $r\sigma_v^2$ and there always exists a pair of finite values of $r\sigma_q^2$ and $r\sigma_v^2$ such that $K_o(t_q) > K_m(t_q)$ for any $t_q \in (0, t^*]$.

A.10. Proof of Proposition 6. Since $D'(t_u) > 0$ for all $t_u > 0$, $t_u = 0$ and $t_q + t_c = t^*$ solves the manager's problem (10). In case of private provision (9), the first-order condition with respect to t_u is $\omega - D'(t_u) - C'(t)$. Since $C'(t) = \max\{\alpha, S'(\underline{t})\}$ the owner will set $t_u > 0$ if and only if $\omega > D'(0) + \max\{\alpha, S'(\underline{t})\}$.

A.11. Numerical example. Assume the functional forms

$$\begin{aligned} C(t) &= (t-2)^8, \\ S(t_c) &= 5\log(t_c+1), \end{aligned}$$

which give

$$\begin{aligned} C'(t) &= 8(t-2)^7, \\ t^* &= 2, \\ S'(t_c) &= \frac{5}{t_c+1}. \end{aligned}$$

Also, set

$$r\sigma_q^2 = 0.45.$$

In the corner solution where $t_q = 0$, we get

$$\begin{aligned} S'(\underline{t}) &= C'(\underline{t}), \\ \frac{5}{\underline{t}+1} &= 8(\underline{t}-2)^7, \\ \underline{t} &= 2.7735. \end{aligned}$$

and so the highest level of α for which $t_q^o(\alpha) = 0$ is

$$\alpha = 8(2.7735 - 2)^7 = 1.3253.$$

In the interior solution, we get

$$\begin{aligned} \alpha &= S'(t_c) = C'(t), \\ \alpha &= \frac{5}{t_c + 1} = 8(t-2)^7, \\ t^o(\alpha) &= \sqrt[7]{\frac{\alpha}{8}} + 2, \\ t_c^o(\alpha) &= \frac{5}{\alpha} - 1, \\ t_q^o(\alpha) &= \sqrt[7]{\frac{\alpha}{8}} + 3 - \frac{5}{\alpha}. \end{aligned}$$

For

$$\alpha \geq \max S'(t_c) = 5$$

we have $t_c = 0$, but this point is never reached in the cases I simulated.

References

Domberger, S and Jensen, P. (1997), "Contracting Out by the Public Sector: Theory, Evidence, Prospects", *Oxford Review of Economic Policy*, Vol. 13, 4, 67-78.

Donahue, John D. (1989), *The privatization decision: public ends, private means*, Basic Books, New York.

Grossman, Sanford J. and Hart, Oliver D. (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy*, 94, Issue 4, 691-719.

Hart, Oliver; Shleifer, Andrei and Vishny, Robert W. (1997), "The Proper Scope of Government: Theory and an Application to Prisons", *Quartely Journal of Economics*, November 1997, Vol. 112, 1127-1161.

Holmström, Bengt and Milgrom, Paul (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics and Organization*, Vol 7., 24-51.

Ohlsson, H. (2003), "Ownership and Production Costs: Choosing between Public Production and Contracting-Out in the Case of Swedish Refuse Collection", *Fiscal Studies*, Vol. 24, 4, 451-476.

Sappington, David E. and Stiglitz, Joseph E. (1987), "Privatization, Information and Incentives", *Journal of Policy Analysis and Management*, 6, 567-582.

Schmidt, Klaus M. (1996), "The Costs and Benefits of Privatization: An Incomplete Contracts Approach", *Journal of Law, Economics and Organization*, Vol. 12, No. 1, 1-24.

Sclar, Elliott D. (2000), *You don't always get what you pay for: the economics of privatization*, Ithaca, New York.

Shleifer, Andrei (1998), "State versus Private Ownership", *Journal of Economic Perspectives*, Fall 1998.

Wilson, James Q. (1989), *Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books.

PAPER 2

When Does Privatization Reduce Costs and Improve Quality? Theory and Evidence from Service Contracting

Erik Lindqvist

ABSTRACT. I develop and test a model of service contracting in residential youth care. Being residual claimants, private providers have strong incentives to keep teenagers in care and cut costs, whereas managers of public facilities lack financial incentives. This difference in basic incentives generates three main predictions. First, when the duration of treatment is insensitive to quality, owners of private facilities must be given rents in order to invest in quality. As a result, private provision is more expensive for high levels of quality, though cheaper for low levels of quality. Second, since public sector managers lack financial incentives, they shun troublesome teenagers. Third, private providers try to prolong the duration of treatment by lying about treatment progress, whereas managers in public facilities tell the truth. The empirical analysis supports the model's predictions. Private facilities have lower per-day cost of treatment for low levels of quality, but higher per-day cost for high levels of quality. Though public facilities generally have a higher level of quality, private facilities are relatively better at treating troublesome teenagers. Treatment periods are much longer under private provision, implying that the average total cost of treatment is twice as high in the private sector. Public provision is superior for non-troublesome teenagers due to lower total cost, but the case of troublesome teenagers is uncertain.

1. Introduction

When is it better for the government to buy a service from a private firm than to produce it itself? This question has been at the center of public debate and academic research for almost three decades. The standard view of economic theory is that private

[0] This work originates from an unpublished paper co-authored by Jens Sandblom, to whom I am very grateful. I am also indebted to Tore Ellingsen for support throughout this project. In addition, I would like to thank Martin Bech Holte, Daniel Ferreira, Guido Friebel, Magnus Johannesson, Henrik Jordahl, Bengt Holmström, Jonathan Levin, Paul Milgrom, Robert Östling, Klaus Schmidt, Per Sonnerby, Giancarlo Spagnolo, Frank Verboven and seminar participants at the CEPR/EBRD Conference on Private and Public Partnerships in London 2007, ESEM 2006, MIT, Norio VI, RIIE (Stockholm), SITE (Stockholm) and SUDSWEc for valuable comments. Bo Vinnerljung and Marie Sallnäs at the Institute for Evidence-Based Social Work Practice (IMS) generously provided me with their data. Finally, the help from Gerd Skans and nine social welfare secretaries have been very important. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

providers are cheaper than public providers, but that they may shirk on aspects of quality that are hard to define in a contract.[1] In this paper, I develop and test a model of privatization of residential youth care based on the multi-tasking model by Holmström and Milgrom (1991). The model generates a number of predictions that differ from the standard view in the privatization literature. First, private facilities are cheaper in terms of per-day cost of treatment for low levels of quality, but more expensive for high levels of quality. Second, public facilities exert less effort on troublesome teenagers, implying that the risk of treatment breakdown in this group is relatively higher under public provision. Third, private providers might prolong the duration of treatment by lying about treatment progress.

The model's predictions are tested on a rich data set of Swedish teenagers placed in youth care during 1991. Using personnel density as a quality measure, I find that private facilities have lower per-day cost of treatment for low levels of quality, but higher per-day cost for high levels of quality. Accordingly, private facilities have lower personnel density, a comparably low level of education among the personnel and a higher frequency of treatment breakdowns. However, as predicted by theory, the risk of a treatment breakdown for teenagers with severe problems is higher in public facilities than in private facilities. While violent and non-violent teenagers experience almost the same probability of treatment breakdown in private facilities, the breakdown frequency is three times higher for violent teenagers than for non-violent teenagers in public facilities. In line with the model's third prediction, the average duration of treatment in private facilities is twice that in public facilities. As the average per-day cost of treatment is only ten percent lower in private facilities, this implies that the total cost of treatment under private provision is almost double that of public provision.

I use data on post-treatment outcomes at the age of 25 to evaluate the long-term effects of facility ownership. For non-troublesome teenagers, there are no significant differences in terms of criminal record, mental health, economic self-reliance and educational attainment between private and public facilities. This suggests that the longer duration of treatment in private facilities compensates for the lower level of quality. However, private facilities are relatively better at treating troublesome teenagers. Violent teenagers in private facilities are relatively less likely to have been convicted in a court of law or to have been imprisoned between the age of 20 and 24 compared to violent teenagers in public facilities. As total cost is twice as high under private provision, public provision of youth care appears to be superior for non-troublesome teenagers, but it is hard to make a definitive assessment for troublesome teenagers.

[1] See, for example, Schmidt (1996), Hart, Shleifer and Vishny (1997) and Shleifer (1998).

Finally, the prevalence of treatment breakdowns and recidivism are lower when the municipality social service plan the duration of treatment in advance, regardless of facility ownership. This suggests that the social services engagement is important for youth care to benefit teenagers.

Residential youth care in Sweden is a market with a large number of private and public facilities which compete on reasonably equal terms. Private and public facilities have the same type of compensation contracts; typically a certain fee per day of treatment. However, as managers of public facilities receive a fixed salary whereas owners of private facilities are residual claimants, private and public provision are in effect associated with very different contracts. As residual claimants, owners have strong incentives both to increase revenue by filling their treatment places and to cut costs. Public facility managers receive a fixed wage and have no financial incentives to attract or keep teenagers at their facility, nor to reduce costs. This difference in basic incentives generates the three main predictions.

First, as contracts on quality are incomplete, private facilities have incentives to invest in quality primarily because it affects their ability to fill their treatment places. This implies that a private provider has a stronger incentive to put effort on quality the more sensitive is the duration of treatment to quality, and the higher is the price of treatment. When social services cannot commit to strong reactions to quality, they thus need to raise prices in order to incentivize private facilities. As a result, private provision will be cheaper for low levels of quality, but more expensive for high levels of quality.

Second, differences in task attractiveness may eliminate the advantage of allocative efficiency under public provision. As wage earners without residual claims, managers of public facilities have no incentive to give extra attention to the most troublesome teenagers. By contrast, owners of private facilities have incentives to keep also the most troublesome teenagers in care.

Third, private firms' stronger incentives to fill their treatment places imply that they try to prolong treatment periods by giving false information on treatment progress. If social services are naive or unwilling to stop treatment prematurely, private providers succeed in prolonging the duration of treatment.

The most extensive previous study on privatization of residential youth care is the paper by Bayer and Pozen (2005) on juvenile correctional facilities in Florida. Using the recidivism rate of young offenders as their measure of quality, they find that private

for-profit facilities have both significantly lower quality and costs than private non-profit and state-owned facilities, whereas county-owned facilities had both lower costs and higher quality than private for-profit facilities.[2]

This paper differs from Bayer and Pozen (2005) in being more closely guided by theory. There are also differences in the empirical results. First, Bayer and Pozen (2005) do not test whether private and public facilities have cost advantages for different levels of quality. Second, even though they suggest that teenagers in private for-profit facilities may serve longer sentences due to owner's incentives to prolong treatment, this hypothesis is not tested explicitly in their paper. Bayer and Pozen (2005) do report one result which indicates that public facilities shun troublesome teenagers: individuals serving especially long sentences had lower recidivism rates when released from a for-profit facility. However, as pointed out by the authors, facilities are able to lengthen or shorten an offender's sentenced commitment period, implying that it is hard to know to what extent sentence length reflects teenager type instead of manager/owner incentives.

The rest of the paper is structured as follows. I describe the market for residential youth care in Sweden in the next section. The model is presented in Section 3 and its predictions discussed in Section 4. I discuss the data set and basic differences between the private and public facilities in Section 5. The results are presented in Section 6. Section 7 concludes the paper. Mathematical proofs, descriptive statistics for the data and some of the robustness checks are provided in Appendix A. Supplementary material are provided in Appendix B.

2. Residential Youth Care in Sweden

Residential care is the most comprehensive measure for youth at risk which the Swedish social services can undertake. In November 2000, about 3,300 Swedish children and teenagers were staying in approximately 500 facilities.[3] Most teenagers are placed in residential care due to their own behavior, such as violent crime, drug addiction or suicidal tendencies. There are two different types of residential youth care facilities in Sweden. The first type, *§12-homes,* are all owned by the government and mainly used for youth convicted for violent crimes. This study focuses on the other type of facility,

[2] In a study on English children's homes, Gibbs and Sinclair (1998) found that residents in private homes were more "difficult" than those in public homes, but the result on quality was ambigous and they did not consider cost. Moreover, there were only eight private homes in their sample, making it difficult to draw any strong conclusions from their study.

[3] Swedish National Audit Office (2002) and National Board of Health and Welfare (2001).

HVB-homes, in which adolescents with a less heavy criminal record are placed. A key difference between these facility types is that only §12-homes can incarcerate teenagers.

The responsibility to act when children have some kind of social problem lies at the municipality level, the lowest tier in Swedish government. It is the municipality social service that acts as buyer in the market for residential youth care. Though each placement must be confirmed by a political committee, the decision to place a teenager in youth care is prepared and implemented by a social welfare secretary, employed by the municipality. At the seller side, public facilities are managed by municipalities or county administrative boards (CABs), whereas private facilities are run both by firms and non-profit organizations.[4] At the time when the teenagers in my data were placed in youth care (1991), there were few formal requirements that hindered entry into the market for youth care.[5] Trade across municipalities and CABs regarding publicly provided youth care is quite limited: In my data, 93.3 % of placements in a municipality facility were done by the municipality owning the facility and 91.5 % of placements in a CAB facility were done by a municipality within the same county.

To get a notion of the social services' working practices, I conducted interviews with nine social welfare secretaries from different municipalities.[6] According to the social welfare secretaries, several considerations influence the choice of facility. First, the facility's treatment program should fit the needs of the teenager. Second, the distance between the facility and the teenager's home should neither be too long (expensive to monitor the facility and harder to integrate a teenager back to a normal life in his or her home environment), nor too short (increases the risk of recidivism and escape). Third, the social services are often unwilling to place teenagers that already know each other in the same facility as this might make it harder for teenagers to change their behavior. Many of the social welfare secretaries also stressed that decisions about residential youth care are often taken under tough time-constraints. Moreover, municipalities often lack information on available facilities and existing treatment options. This is a particular problem for small municipalities that place few teenagers.[7] In the absence

[4] The counties constitute the second tier in Swedish government, in between the State and the municipalities.

[5] See Swedish National Audit Office (2002) for a discussion of entry requirements in youth care. A set of stricter requirements for starting a youth care facility were introduced on January 1st 2002. For example, managers at private facilities now have to hold a university degree in a field relevant for youth care, such as psychology or social work.

[6] Eight interviews were undertaken in 2005 and one interview in 2002.

[7] According to Sallnäs (2005), 87 percent of 97 interviewed private facility managers said that the municipalities' most important source of information about their facility was their previous experiences. A quote from the interviews also illustrates this point: "In our municipality, one teenager may get a serious problem of drug addiction every five years or so. We then have to learn everything about the market for residential youth care all over again."

of personal experience, most of the interviewed social welfare secretaries said that they collect references on facility performance from other municipalities.

There is no standardized contracting procedure between buyers and sellers in the market for residential youth care, and it is up to the municipality and facility to negotiate the contractual terms, including the price of treatment.[8] However, Swedish law limits the set of tasks that a municipality can contract out. For example, it is not possible for a municipality to let a private facility take final responsibility for the treatment. A "standard" contract stipulates what kind of treatment the teenager should undergo and how it is supposed to be documented, how contacts between the social services and the facility should proceed, the price of treatment and period of notice. Facilities are typically compensated per day of treatment. Most contracts entail few direct incentives for quality provision.[9]

Public facilities can be separate legal entities (i.e., firms) or separate units within the municipality or CAB administration. When youth care is provided in-house, the social services are still charged a price for treatment, which is set internally in case of a municipality facility, or by the CAB in case of a CAB facility.[10] In their discussion of the cost of treatment in youth care, Vinnerljung et al. (2001), whose data I use, treat prices in public facilities as reflecting true costs. Nevertheless, the possibility that internal prices do not perfectly reflect costs cannot be ruled out. For example, public HVB-homes might be charged below-market rents for their facilities or receive loans on favorable conditions.

To the best of my knowledge, managers of public facilities all get a fixed salary and there are no public managers with explicit incentive contracts. At a direct question, most interviewed social secretaries said that the main difference between private

[8] The discussion about contracting procedures is partly based on a number of actual contracts I was able to study in connection with the interviews. It is uncertain how accurate these contracts are as indicators of a typical contract in 1991, the year the teenagers in our data set were first placed in residential youth care. However, virtually all of the interviewed social welfare secretaries said that contracting procedures have become more rigorous since the early 1990's. As a concrete example of this, many social welfare secretaries mentioned that the period of notice is shorter today than in the early 1990's. Another example is that the municipalities since January 1st 2002 are required by Swedish law to establish a plan for the treatment which has to be signed by the municipality, the teenager's parents and the facility. In this way, the state is thus forcing the municipalities to sign more explicit and elaborate contracts.

[9] However, some contracts stipulate a direct cut in treatment fees should the teenager run away or the facility not undertake the treatment stated in the contract.

[10] The division of the municipality organization into separate units responsible for covering costs and the use of an internal price system became common within Swedish municipalities during the 1980's (Haglund et al., 1993). In the data I analyze, 92.5 % of files for teenagers in public facilities contained data on costs, compared to 87.2 % of files for teenagers in private facilities. There is thus strong reason to believe the missing data on costs in public facilities reflect incomplete documentation, not the lack of a system of internal prices.

and public youth care facilities is the private facilities' stronger incentive to fill their treatment places.

One likely reason for why municipalities and facilities do not sign explicit contracts on quality is that many aspects of residential care are inherently difficult to quantify and might even be subject to secrecy (an example is therapy). There is also considerable ambiguity as to exactly what constitutes good youth care; there is an abundance of different theories on how problematic teenagers should best be treated, with relatively little agreement on basic principles (Sallnäs 2000).[11] In addition, facilities are typically situated at some distance from the municipality center, making monitoring visits costly. Another difficulty stressed by several of the interviewed social welfare secretaries is that treatment quality is sensitive to changes in the personnel force.[12]

However, there are also indications that the social services put little effort into writing contracts and monitoring facility performance. According to Sallnäs (2005), 60 percent of 97 interviewed managers at private facilities said that the social services "rarely" or "never" asked for evaluations of treatment quality when placing a teenager at their facility. The Swedish National Audit Office (2002) argues that the municipalities' and CABs' lack of adequate monitoring of quality is a major problem of Swedish residential youth care. In a extensive survey of the research on youth care, Andreassen (2003) concludes that a large fraction of residential youth care is not undertaken according to the established principles of effective treatment.

One likely explanation for the low emphasis put on monitoring facility performance is the weak incentives of municipality bureaucrats. Alternatively, some social welfare secretaries might not believe that explicit contracts and monitoring visits are important. Few social workers have received training in economics or business administration and the literature on youth care in Sweden does not contain much discussion about incentives and contracts.[13] Social workers have to learn about good contracting procedures by experience, but since many municipalities are small and place few teenagers, this could take a long time. During the interviews, many of the social welfare secretaries emphasized that they follow up their placements more closely today than in the

[11] A common view expressed both in the literature (e.g. Sallnäs 2000) and in the interviews is that *which* method is used is not as important as the fact that *some* method is used. In the words of one of the interviewed social secretaries: "The personnel of the facility need something to hold on to. A theory can provide that."

[12] One of the interviewed social welfare secretaries said that: "It takes a long time to get to know a facility, and once you do it's not the same anymore." The same social secretary also said that they deliberately choose to work with few facilities in order to get more accurate information, thereby sacrificing a larger set of treatment options.

[13] For example, the National Board of Health and Welfare issued three reports on privatization of residential youth care in the early 1990's (1990, 1991 and 1994), but none of them has an explicit discussion of facility manager incentives.

early 1990's, partly because the law now obliges them to do so and partly because of an increase in experience and professionalism.

The next section develops a model of residential youth care that captures the institutional features described above. The model generates testable predictions of differences between private and public providers that I then take to the data.

3. A Model of Residential Youth Care

The model is based on the multi-tasking model by Holmström and Milgrom (1991). Consider a municipality social service that must buy residential youth care for a certain teenager. Let the severity of a teenager's problem be indexed by v, where higher v implies a more troublesome teenager. The agent is a manager in case youth care is organized in-house or an owner of a private facility. Let the superscript o denote the owner and the superscript m the manager. For simplicity, there is only one period in the model.

The agent can invest effort in three different tasks, the effects of which last the entire treatment period. The first task (t_c) reduces the cost of treatment and the second task (t_q) improves general treatment quality. In addition, there is a second aspect of quality (t_v) that only pertains to troublesome teenagers, i.e., those for which $v > 0$. The three tasks differ in their inherent attractiveness. Total effort on cost reductions and general quality, $t = t_q + t_c$, entails a cost $C(t)$ for the service provider. The cost function, $C(t)$, is minimized at $t = t^* > 0$, i.e., people are assumed to exert some effort even in the absence of financial incentives. In contrast, treating troublesome teenagers entails a cost $D(t_v)$ which is increasing in t_v for all $t_v \geq 0$. Both $C(t)$ and $D(t_v)$ are strictly convex and twice continuously differentiable in their respective arguments.[14] I normalize the cost of effort so that $C(t^*) = D(0) = 0$. The service provider's utility is $u(x, t, t_v) = x - C(t) - D(t_v)$, where x is the financial compensation. The service provider is assumed to get a utility of zero in his best outside option.

The cost reduction is given by a function $S(t_c)$ which is increasing, strictly concave and twice continuously differentiable in t_c.[15] Total cost is thus decreasing in $S(t_c)$. To ensure non-negative prices, I assume that each treatment place must cover a fixed cost $c > S(\underline{t})$ where $\underline{t}$ is such that $C(\underline{t}) = S(\underline{t})$.

The general quality of treatment is influenced by aspects beyond the facility's control, such as peer effects within the group of teenagers. Formally, let $q = t_q + \varepsilon$ denote the quality outcome where $\varepsilon \sim U(-k, k)$. Total treatment quality for a teenager is

[14] To simplify the analysis, I also assume that $\lim_{t \to \bar{t}} C'(t) = \infty$, $C'''(t) > 0$ and $\lim_{t \to \bar{t}} C''(t) = \infty$ for some finite $\bar{t}$.

[15] To simplify the analysis, I also assume that $S'''(t_c) > 0$.

given by $q + \min\{t_v - v, 0\}$.[16] That is, I assume that troublesome teenagers need extra care and attention in addition to the efforts put down to provide all teenagers with a high treatment quality.

For each teenager in care, the social service has a preference for the duration of treatment, T^*.[17] When treatment has lasted less than T^* days, every day spent at the facility gives a benefit $B(t_q, \psi)$ which is increasing and strictly concave in t_q with some finite upper bound for $\psi > 0$. The parameter ψ denotes the public agency's valuation of youth care quality. For all $t_q > 0$, $B(t_q, \psi)$ is increasing in ψ with $\lim_{\psi \to \infty} B(t_q, \psi) = \infty$, $B(t_q, 0) = a$ for all t_q, and $B_{t_q}(t_q, \psi)$ is increasing in ψ.[18] The benefit from an additional day of treatment when treatment has lasted more than T^* days is equal to zero.

Let p denote the per day cost of treatment and T the total duration of treatment in days. The social service's surplus S from treatment is then given by

$$S = \begin{cases} (B(t_q, \psi) - p)\, T & \text{for } T \leq T^*; \\ (B(t_q, \psi) - p)\, T^* - p(T - T^*) & \text{for } T > T^*. \end{cases}$$

An important feature of residential youth care is that treatment often breaks down before it has been intended to end. For example, the teenager might run away from the facility. Let $\underline{q}$ denote the minimum level of quality required for treatment not to break down, i.e., treatment breaks down if $q + \min\{t_v - v, 0\} < \underline{q}$. Unless $t_v = v$, teenagers with severe problems are thus more likely to experience a treatment breakdown. Whereas treatment breakdown is a direct function of quality, I assume that the effect of quality on duration of treatment is determined by the social service's reaction to low or high levels of quality.[19] In terms of the model, T^* is increasing in

[16] I thus ignore uncertainty in the case of t_u. Introducing uncertainty here as well would complicate the model without changing the results in any substantial way.

[17] T^* need not represent the societal optimum. For example, politicians and social welfare secretaries may have wrong beliefs about the effect of treatment. Alternatively, some municipalities may have a self-interest in keeping teenagers at the facility. There is some indication in the literature on youth care that this is an important problem. Some writers and researchers on residential youth care have alleged that youth care *primarily* is a means for society of getting rid of "disturbing elements" (see Sallnäs 2000, p. 107 for a discussion). Andreassen (2003) claims that current research indicate that the optimal duration of treatment is about six months, a considerably shorter duration of treatment than that experienced by the majority of teenagers in my data. My perspective here is thus strictly positive and I am agnostic about the impact on post-treatment outcomes of treating teenagers for a certain period of time.

[18] I assume that a is so large that the social services want to buy youth care as long as long as $E[T^*] > 0$.

[19] This is of course a simplification. In reality, teenagers affect the duration of treatment by running away from the facility and treatment breakdowns occur because the social service reacts to a low level of quality. However, it is not unreasonable to assume that the social service's preferences are more important for the duration of treatment than for the probability of a treatment breakdown. Treatment breakdowns are rather seldom initiated by the social service (14 % of treatment breakdowns) and,

$\alpha\left[q+\min\left\{t_v-v,0\right\}-\underline{q}\right]$.[20] I assume that $\alpha \geq 0$ for all social services, implying that the duration of treatment is weakly increasing in quality. There are three reasons for this assumption. First, the pressure of social welfare secretaries to act to shorten the duration of treatment is arguably stronger when quality is bad than when it is good. Second, many placements end in a treatment breakdown, the risk of which is probably affected by quality. Third, though there is only one placement in the model, the interviews indicated that social welfare secretaries are more likely to send teenagers to a facility in the future if quality is high. Assuming that $\alpha > 0$ is therefore a reduced form of incorporating facilities' concern for a good reputation.

Another feature of residential youth care is that the effect of treatment is hard to predict. Once treatment has started, facilities receive private information on treatment progress. Let $D \in \{-1,0,1\}$ denote the true status on treatment progress which is observable to the facility but not to the social services.[21] $D=-1$ implies that treatment should be shortened by $\gamma > 0$ days and $D=1$ that treatment should be prolonged by γ days. The optimal duration of treatment is then given by

$$T^* = \alpha\left[q+\min\left\{t_v-v,0\right\}-\underline{q}\right]+\beta+\gamma D,$$

where β denotes the duration of treatment which is independent of quality and information about treatment progress.

Facilities send an unverifiable signal $I \in \{-1,0,1\}$ to the social service with the same interpretation as D. The social services then instruct the facility to take an action $A \in \{-1,0,1\}$, i.e., they decide whether or not the duration of treatment should be shortened or prolonged by γ days after observing the facility's signal.[22] There are two types of social welfare secretaries. *Sophisticated* social welfare secretaries have rational expectations whereas *naive* social welfare secretaries take the facility's signal at face

when treatment breakdown do not occur, the social services have the final say over the duration of treatment.

[20] I thus separate the municipalities' willingness to pay for quality (ψ), and the effect of quality on duration of treatment (α). This assumption simplifies the proofs in the paper, but can also be motivated on other grounds. For example, reacting to observed quality requires effort on the social secretary's behalf (for example, if q is low the social welfare secretary may have to find another treatment place for the teenager in question). If the social secretary care for quality subject to the constraint that he doesn't have to work hard, α may be low while ψ is high. Another argument for separating ψ and α is that the willingness to pay for quality is limited by the social services budget, which is decided upon by politicians, whereas actions are taken by employed bureaucrats. Hence, whereas the politicians' preferences determine ψ, it is the social welfare secretary's preferences that determine α.

[21] The "true" duration of treatment here refers to the duration of treatment the social services would choose if they knew about treatment progress, not to the societal optimum.

[22] Given the formulation of S, this formulation of the action space is not restrictive. That is, it is never optimal for a social service to reduce or prolong the duration of treatment by anything else that γ.

value. I assume that a positive proportion of social welfare secretaries are naive.[23] For simplicity, the effects of quality and information on treatment progress are additive, implying that the decision whether treatment should be prolonged or shortened does not affect the risk of treatment breakdown. After observing q and I, the social service will set

$$T = \begin{cases} \alpha\left(q - \underline{q} + \min\left\{t_v - v, 0\right\}\right) + \beta + \gamma I & \text{if naive and} \\ \alpha\left(q - \underline{q} + \min\left\{t_v - v, 0\right\}\right) + \beta + \gamma \arg\max_A E\left[S \,|I, \text{Owner}\right] & \text{if sophisticated.} \end{cases}$$

I assume that the social service cannot sign contracts with explicit incentives for quality provision or cost reductions ex ante, and that facilities cannot commit to a certain level of quality.[24] Since price is fixed at the starting day of treatment, the duration of treatment (T) determines a facility's revenue once treatment has started.[25] The owner of the private facility is the residual claimant to the firm's profits whereas managers of public facilities just receive a fixed wage, w. The wage w is independent of the duration of treatment, i.e., employment contracts are not conditioned on the number of teenagers in care. I thus model wage contracts as a lump-sum transfer. As each facility has several placements, w is not equal to the total wage cost, but denotes the share of wage costs that a given placement must bear.

The assumption that public facility managers only receive a fixed wage reflects the institutional features in the market for residential youth care, but it can also be given a theoretical foundation. If the government retains ownership, contracts on profit sharing could be manipulated ex post. Though it might be possible to write contracts that restrict the government's scope for opportunistic behavior, this is costly. Moreover, the task of writing contracts falls on municipality bureaucrats with weak incentives to search for innovative contracting solutions.

It is hard to know what are the incentives for managers in private non-profit facilities. For example, it is not obvious that non-profit organizations are not motivated by profit. I also lack information on the ability of non-profit facility managers to influence their own compensation. Given these difficulties, I focus on the clearer distinction

[23] Due to the weak incentives for social service bureaucrats, a small cost of effort associated with learning about the true model is sufficient for a "rational" social welfare secretary to act in a naive fashion.

[24] This may seem like stark assumptions. However, for the argument in this paper to be valid, it does not matter whether the social welfare secretaries *cannot* sign more elaborate contracts, or whether they *will not* sign such contracts due to weak incentives.

[25] Technically, a sufficiently large k will imply that values below zero is in the support of T^* and T. This feature could be ruled out by making specific assumptions regarding the size of k, or by truncating T. Alternatively, one could think of $T < 0$ as a reduced form of a negative reputation effect. As it is not important for the results derived in the paper, I will henceforth disregard this issue.

between private for-profit and public facilities, but control for non-profit ownership in the empirical analysis.

Given the assumptions above, the financial return for owners, x^o, is $pT + S(t_c) - c$, and so

$$u^o = pT + S(t_c) - c - C(t) - D(t_v).$$

As $x^m = w$, we get

$$u^m = w - C(t) - D(t_v)$$

for public facility managers.

3.1. Public Facility. The manager's maximization problem with respect to effort is

$$\max_{\{t_q, t_c, t_v\}} \quad w - C(t) - D(t_v). \tag{3.1}$$

Since the manager has no financial incentives, he just minimizes his cost of effort, implying that $t_v = 0$ and $t_q + t_c = t^*$ where $t^* = \arg\min C(t)$. As the manager is indifferent between t_q and t_c, I assume that he divides his effort between the two tasks according to the wishes of the social service.[26] Similarly, I assume that the manager reveals the truthful state of treatment progress as he is indifferent to the duration of treatment.[27]

As the residual claimant, the social service receives the full share of the cost savings and has to pay the fixed cost c. I assume that the prices paid under public provision reflect expected costs. Since the social service has a symmetric prior distribution on D, the social service's maximization problem can be formulated as

$$\begin{array}{ll} \max\limits_{\{t_q, w\}} & B(t_q, \psi)\left(\alpha t_q - \alpha \underline{q} + \beta\right) + S(t^* - t_q) - w - c, \\ \text{s.t.} & w \geq 0, \\ & t_q \in [0, t^*]. \end{array}$$

It is straightforward to see that the social service always sets $w^m = 0$. The optimal level of effort on quality $\left(t_q^m\right)$ in the interior solution is such that $B_{t_q}\left(t_q^m, \psi\right)\left(\alpha t_q^m - \alpha \underline{q} + \beta\right) + B\left(t_q^m, \psi\right)\alpha = S'\left(t^* - t_q^m\right)$. That is, the social service instructs the manager to set his effort on quality such that the marginal benefit of quality equals the marginal cost reduction.[28]

[26] We could think of this assumptions as the manager having lexicographic preferences where his first priority is to maximize his own utility and his second priority to act according to the wishes of the social service.

[27] Assuming a direct concern for truth-telling rules out nonsensical equilibria that may arise if the manager instead cared about the social service's reaction to his information.

[28] The complete solution is provided in the Appendix.

As the manager speaks the truth about treatment progress, the social service follows the managers recommendation.

3.2. Private Facility. Let $\alpha p = \widetilde{\alpha}$, $\gamma p = \widetilde{\gamma}$ and $\beta p = \widetilde{\beta}$. The owner's maximization problem with respect to effort is then

$$\max_{\{t_q, t_c, t_v, I\}} \begin{array}{l} \widetilde{\alpha} t_q - \widetilde{\alpha} \underline{q} + \widetilde{\alpha} \min\{t_v - v, 0\} + \widetilde{\beta} + S(t_c) \\ -C(t) - D(t_v) - c. \end{array} \tag{3.2}$$

If $S'(\underline{t}) < \widetilde{\alpha} < S'(0)$, where $\underline{t}$ is such that $S'(\underline{t}) = C'(\underline{t})$, the owner sets t_q and t_c such that

$$\widetilde{\alpha} = S'(t_c) = C'(t).$$

When $v > 0$ and $\widetilde{\alpha} < D'(v)$, the owner sets t_v such that

$$\widetilde{\alpha} = D'(t_v).$$

Due to the strict concavity of $S(t_c)$ and strict convexity of $C(t)$ and $D(t_v)$, these interior solutions are unique. The complete solution to the owner's problem (including corner solutions) is provided in the Appendix. In the interior solution, the owner's effort on cost reductions is continuously falling in $\widetilde{\alpha}$. That is, as incentives for quality gets stronger, the owner shifts his effort from reducing costs to increasing quality. As shown in the Appendix, the owner's effort on quality is continuously increasing in $\widetilde{\alpha}$ for $\widetilde{\alpha} > S'(\underline{t})$. Similarly, the owner's effort on troublesome teenagers is continuously increasing in $\widetilde{\alpha}$ up to $t_v = v$ for $\widetilde{\alpha} > D'(0)$. Hence, for each α and $t_q > 0$, there is a unique price such that the owner of the firm sets $t_q^o = t_q$, and similarly for t_v above zero and below v. By raising the price, the teenager becomes more valuable to the facility, implying that incentives for both quality provision and effort on troublesome teenagers are strenghtened.

As the owner only cares about prolonging treatment, the signal is completely uninformative and not believed by a sophisticated social service. However, since there is a positive probability that the social welfare secretary is naive, the owner always sets $I = 1$.

As the social service cannot contract on the duration of treatment, the only contractual parameter is the price of treatment. For $p \geq S(\underline{t})/\alpha$, quality is increasing in p, implying that both the value of each day in care and the duration of treatment is increasing in p. It is straigthforward to show that the owner's utility is strictly increasing in the price of treatment, and that there is a unique price denoted by p^* for each

set of parameters $(\alpha, \beta, v, \gamma)$ and reactions to information on treatment progress (A) such that $u^o = 0$.

A social services' maximization problem under private provision is different depending on whether he is sophisticated or naive. A naive social welfare secretary expects to receive correct information on treatment progress. Since the social welfare secretary has a symmetric prior distribution on D, the maximization problem can then be written as

$$\begin{array}{ll} \max\limits_{p} & \left(B\left(t_q\left(\widetilde{\alpha}\right), \psi\right) - p\right)\left(\alpha t_q\left(\widetilde{\alpha}\right) + \alpha \min\left\{t_v\left(\widetilde{\alpha}\right) - v, 0\right\} - \alpha \underline{q} + \beta\right) \\ \text{s.t.} & p \geq p^*. \end{array} \tag{3.3}$$

If a social welfare secretary knows that he will not receive accurate information on treatment progress, he simultaneously decides what price to pay and whether to plan for shortened or prolonged duration of treatment. The formal solution to the sophisticated social service's maximization problem is provided in the Appendix.

3.3. Results. The highest possible level of quality under public provision is t^*. In that case, the social service orders the manager of the public facility to spend all his efforts on increasing quality. Under private provision, the teenager in question becomes more valuable to the owner when price goes up, implying that the incentive for quality increases. For every $\alpha > 0$, there exists a price such that quality is higher under private provision than under public provision.

PROPOSITION 1. *If $\alpha > 0$, the highest feasible level of quality is always higher under private provision.*

Proofs are provided in the Appendix.

An increase in price also increases the owners incentive to put effort on troublesome teenagers. In contrast, managers of public facilities have no incentive to retain teenagers and therefore put in no such extra effort.

PROPOSITION 2. *Public facilities exert no extra effort on troublesome teenagers, but private firms do so if $\widetilde{\alpha}$ is high enough.*

In analyzing how costs for a given level of quality depend on α, I make the additional assumption that $\beta = \beta' - \alpha\left(E\left[t_q\right] - \underline{q}\right)$. This assumption implies that the expected duration of treatment for non-troublesome teenagers is independent of α, allowing us to focus on the cost of providing incentives in the private sector.[29]

[29] The assumption that $\beta = \beta' - \alpha\left(E\left[t_q\right] - \underline{q}\right)$ should not be interpreted as suggesting that municipalities preferences depend on the level of quality. Rather, I am making a different assumption on the distribution of α and β in the population of municipalities for each level of quality that I

When α is low, private provision is cheaper for no effort on quality due to stronger incentives to reduce costs. To see this, note that the cost reduction is equal to $S(t^*)$ under public provision and no effort on quality. Since $C'(t^*) = 0 < S'(t^*)$, this is a suboptimally low level of effort. In contrast, the owner of the private facility receives the full amount of cost savings and sets $t_c^o = \underline{t} > t^*$, which is the efficient level of effort. However, a low α also implies that the social service must set a high price in order to incentivize the owner to put effort on quality, implying that the owner will earn rents and that private provision is more expensive for high levels of quality.[30] When α is high, price need not be high in order to incentivize the owner, implying that private provision is cheaper than public provision for high levels of quality due to stronger incentives for cost reductions. However, as the price that gives a certain level of quality is falling in α, there is a threshold on α for all levels of quality above which private provision is not feasible.

PROPOSITION 3. *Suppose that* $\beta = \beta' - \alpha\left(E[t_q] - \underline{q}\right)$, *and that the social services do not shorten the duration of treatment under private provision. Then there exists a finite threshold* $\widehat{\alpha}_1$ *such that private provision is cheaper for no effort on quality for all* $\alpha \leq \widehat{\alpha}_1$, *whereas no effort on quality is not feasible under private provision for* $\alpha > \widehat{\alpha}_1$. *Second, for all positive levels of quality not above the maximum level of quality under public provision,* t^*, *there exist two thresholds,* $\overline{\alpha}_2$ *and* $\widehat{\alpha}_2$, *such that private provision is feasible but more expensive for all* $\alpha \in (0, \overline{\alpha}_2)$, *feasible and cheaper for all* $\alpha \in (\overline{\alpha}_2, \widehat{\alpha}_2]$, *and not feasible for* $\alpha > \widehat{\alpha}_2$.

Proposition 3 might not hold when the social services shorten the duration of treatment in private facilties due to the lack of accurate information on treatment progress. If γ is sufficiently large, the need for private providers to cover their costs in a shorter duration of time dominates private providers stronger incentive to cut costs, implying that private provision is more expensive for all levels of quality regardless of α. If private facilities succeed in prolonging treatment, private providers become cheaper for

consider. For a fixed distribution of α and β, α is correlated with the duration of treatment, and the correlation varies with the level of quality. This complicates the analysis as cost in the public sector and reservation prices among private facilities are decreasing in the expected duration of treatment. However, though it is certain that α will be correlated with the duration of treatment, we cannot say whether the correlation is positive or negative for a given level of quality unless we make specific assumptions on the correlation between α and β, and the size of $\underline{q}$.

[30] When contracts entail rents, private facilities are willing to pay the social service a sum corresponding to the entire rent for obtaining a contract. However, payments for contracts would probably be politically controversial as it implies that municipalities "sell" teenagers to facilities. In addition, it would invite moral hazard on the municipality's part and require that facilities are not constrained by limited liability. For these reasons, I rule out such transfers by assumption. Still, even if facilities cannot pay municipalities for lucrative contracts, they might spend resources trying to win them. For example, facilities may invest in marketing.

no effort on quality. In addition, it increases the cutoffs $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ for each level of quality, implying that each level of quality is feasible for a larger range of values of α and lower prices, though the price for each given α is the same.

Though Proposition 3 holds also when teenagers are troublesome, the fact that a teenager is troublesome will raise prices for all levels of quality under public provision (due to a shorter expected duration of treatment). Troublesome teenagers also raise prices for no effort on quality under private provision. In addition, it decreases $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ for each level of quality, implying that each level of quality is feasible for a smaller range of values of α and higher prices, though the price for each given α is the same.

As argued above, public providers always provide correct information on treatment progress, whereas private providers always try to prolong treatment.[31] Private providers succeed in prolonging treatment if the social service is naive. The model is not sufficiently specified to derive results regarding the behavior of sophisticated social services. For example, it is not obvious that the incentive for prolonging treatment is increasing in the valuation of quality. On the one hand, the per-day benefit of treatment, $B(t_q, \psi) - p$, is increasing in the valuation of quality. On the other hand, as optimal prices increase in the valuation of quality, the cost of prolonging treatment beyond T^* also increases.

PROPOSITION 4. *Under public provision, the social welfare secretary expects and receives accurate information on treatment progress regardless of whether he is sophisticated or naive. A naive social service will prolong treatment under private provision.*

4. Predictions

There are a couple of reasons to expect that α is relatively low. First, many municipalities seldom place teenagers in care, implying that reputation effects are not likely to be strong. Second, social welfare secretaries have weak incentives to take action when quality is low. Third, as argued by some of the interviewed social welfare secretaries, teenagers tend to get rooted at facilities, implying that it is not self-evident that it is optimal to relocate teenagers even if quality is low. Moreover, the actual incentives faced by owners do not only depend on the social services' reactions to quality, but also on how easy it is for facilities to fill their treatment places with new teenagers. As the number of teenagers in HVB-homes expanded rapidly during the

[31] I assume that there is no fixed cost of treating troublesome teenagers. That is, managers of public facilities are in fact indifferent between treating troublesome and non-troublesome teenagers. If they are not, and wages don't adjust to changes in the composition of teenagers, managers of private facilities have an incentive to send the "shorten" signal upon receiving a troublesome teenager.

1990's (Sallnäs, 2000), this was probably not very hard. Hence, as shown in Proposition 3, private provision may be more expensive for high levels of quality because the social service's raise prices to incentivize owners.

When α is low, prices in the private sector for high levels of quality will not be affected by whether treatment is shortened or prolonged, or whether teenagers are troublesome or not. The reason is that as prices need to be set high anyway to provide owners with incentives for quality, the only impact of changes in the expected duration of treatment is the extent of owners' rents. In contrast, teenagers with a long expected duration of treatment should have lower prices for all levels of quality under public provision.

P1: Private facilities are cheaper for low levels of quality but more expensive for high levels of quality.

The probability of treatment breakdown for non-troublesome teenagers is

$$P\left(q < \underline{q}\right) = P\left(t_q + \varepsilon < \underline{q}\right).$$

If $\underline{q}$ is the same for all teenagers, the probability of treatment breakdown for non-troublesome teenagers thus depends on the effort on quality. As shown above, it is not obvious from a theoretical viewpoint whether quality will be highest under private or public provision. Yet, as we will see, the private facilities in my data have significantly lower personnel density, a personnel force with a comparably low level of education, and there is no evidence of the non-monotonicity in quality suggested by theory. As an empirical matter, we can therefore assume that $t_q^o < t_q^m$, implying that the frequency of treatment breakdowns for non-troublesome teenagers should be higher under private provision.

P2: Non-troublesome teenagers have a higher risk of treatment breakdown in private facilities than in public facilities.

The probability of treatment breakdown for a troublesome teenagers is

$$P\left(q + t_v - v < \underline{q}\right) = P\left(t_q + \varepsilon + t_v - v < \underline{q}\right)$$

where $t_v \leq u$. Since $t_v^m = 0$, but $t_v^o > 0$ when $\widetilde{\alpha}$ is sufficiently high, it is uncertain from a theoretical perspective whether the breakdown frequency for troublesome teenagers will be higher in private or in public facilities. However, the severity of a teenager's problems should be a stronger predictor for treatment breakdowns in public facilities.

P3: Troublesome teenagers have a relatively higher risk of treatment breakdown in public facilities.

The expected duration of treatment under public ownership is

$$E\left(T^m\right) = \alpha\left(t_q - \underline{q} - v\right) + \beta.$$

For a naive social service, the expected duration of treatment under private provision is

$$E\left(T^o\right)_{Naive} = \alpha\left(t_q - \underline{q} + \min\left\{t_v - v, 0\right\}\right) + \beta + \gamma.$$

Controlling for effort on general quality and troublesome teenagers (t_q and t_v), teenager characteristics (v) and municipality preferences (α, β and $\underline{q}$), private provision should prolong treatment by γ days due to false information on treatment progress when social services are naive. This effect will not be foreseen and the actual duration of treatment should therefore exceed planned duration of treatment.

When the social service is sophisticated, the expected duration of treatment is given by

$$E\left(T^o\right)_{Sophisticated} = \alpha\left(t_q - \underline{q} + \min\left\{t_v - v, 0\right\}\right) + \beta + \gamma \arg\max_A E\left[S\right].$$

In this case, planned duration of treatment will equal actual duration of treatment in expectation. As argued above, the model does not specify how sophisticated social services will behave regarding the duration of treatment. However, there are a couple of factors not included in the model which suggest that sophisticated social services may prolong treatment also in cases where they know that this is not optimal. First, social welfare secretaries that do not prolong treatment may have to engage in unpleasant conflicts with owners of private facilities. Second, as private providers are likely to muster any kind of evidence which suggests that treatment should be prolonged, social welfare secretaries might have to work hard in order to provide counter-arguments for why treatment should not continue. Third, critics of the social services tend to focus on whether youth at risk are provided with care, not on the actual effects of treatment.[32]

The model does not predict whether post-treatment outcomes will be better or worse under private provision. Even though the general level of quality is lower under private provision, teenagers in private facilities could have better final outcomes if private provision increases the duration of treatment and the extra time spent in the facility has a beneficial effect. However, we would expect troublesome teenagers to have relatively better outcomes in private facilities.

[32] See Vinnerljung et al. (2001) for a discussion of this issue.

P4: Troublesome teenagers have relatively better post-treatment outcomes if treated in private facilities.

Finally, the social welfare secretary's effort might be important for outcomes in ways not captured by the model. For example, social welfare secretaries may provide the facility personnel with information about the teenager's problem background and specific needs of treatment. Engaged social welfare secretaries are also likely to put down more effort on helping the teenager after treatment has ended.

P5: Engaged social welfare secretaries reduce the risk of treatment breakdown and improve post-treatment outcomes.

Given that α is low and that ψ is not very high, we can establish a number of results regarding the selection of teenagers into private or public facilities. When quality is considered unimportant, naive social services will prefer private provision due to lower cost per day. As sophisticated social services take private facilities distortion of information on treatment progress into account, it is not certain that they will buy low-quality care from private facilities. When quality is important, both naive and sophisticated social services will consider public provision superior for non-troublesome teenagers due to the lower per-day cost for high levels of quality. Selection remains unclear when teenagers are troublesome and quality important. On the one hand, public facilities are cheaper for high levels of general quality and provide correct information. On the other hand, private facilities put more effort on troublesome teenagers.

5. Data

I use a data set compiled by Vinnerljung, Sallnäs and Kyhle-Westermark (2001) at the Institute for Evidence-Based Social Work Practice (IMS), National Board of Health and Welfare (NBHW) in Sweden. The data is based on the files of all Swedish adolescents (13-16 years of age) who were placed in a HVB-home during 1991, with the exception of teenagers who were only placed temporarily in youth care or for the sole purpose of having their treatment needs evaluated before assigned to their final placement.[33] There are some missing observations where the files could not be found

[33] Since some facilities did not receive any teenagers in 1991, this data set only covers a subsample of all HVB-homes in Sweden that were active in 1991.

or were impossible to interpret.[34] I also exclude one facility from the sample where treatment took the form of long, large-scale sailing trips. The teenagers were followed as long as they were subject to residential care, or until their 18th birthday. In addition, the data set contains information on post-treatment outcomes at the age of 25.[35] In total, the data set consists of 357 placements of 336 different teenagers in 173 facilities (186 placements in 63 public facilities and 171 placements in 111 private facilities).[36] The number of observations from a single facility varies from 1 to 10. Using NBHW registers, I have separated privately owned facilities run by firms (78) from those run by non-profit organizations (20). There are 13 cases where I could not determine whether a facility should be considered "for-profit" or "non-profit".[37] The public facilities in the sample are owned by CABs (43) or municipalities (16). There are 3 public facilities in the sample that could not be classified either as a CAB or a municipality facility.[38] Summary statistics for the variables in the data are listed in Table A1 in the Appendix.

The teenagers' differ in terms of treatment history and the reason for placement. Sixty-one percent of the teenagers in the data were placed in youth care due to their own behavior. The second and third most common motives for placement were relational problems (39 percent) and lack of care and attention (27 percent).[39] There is a large variation in the teenagers' problems: 22 percent had psychological problems, 34 percent were addicted to drugs, 40 percent had committed crimes and 19 percent were considered violent. Teenagers often had multiple problems. For example, 67 percent of violent teenagers were criminal compared to 34 percent of non-violent teenagers.

[34] This was the case for about one in three placements from Stockholm and Malmö, Sweden's largest and third largest city, but only for about one in thirty placements from the rest of the country. According to Vinnerljung et al (2001), there is nothing that indicates that the missing files were concentrated on a certain group of children. Yet if Stockholm and Malmö differ systematically from the rest of the country in terms of the teenagers they place in residential care, the sample will not be fully representative.

[35] The data on post-treatment outcomes come from various sources of register data collected by different government agencies. The data have been compiled by Bo Vinnerljung and Marie Sallnäs at IMS who generously provided it to me.

[36] As my interest is in the behavior of facilities, I use placements as my unit of analysis. This raises two different issues. First, as some teenagers experienced more than one placement, the sample is not perfectly representative with respect to the set teenagers that were ever placed in youth care. Second, placements for the same teenager might not be independent observations. In order to check the first problem, I have run the main regressions reported below excluding all placements in HVB-care expect the first for every teenager. To deal with the second problem, I run the same regressions with standard errors clustered at the teenager level. The results do not change substantially in any of these cases and are available upon request.

[37] There were 107 placements in private firms, 51 in non-profit organizations and 14 in facilities that could not be classified.

[38] There were 134 placements in CAB facilities, 48 in municipality facilities and 5 in facilities that could not be classified.

[39] There can be more than one motive for placing a teenager in youth care.

Boys were more likely to be placed in youth care due to their own behavior than girls (70 percent compared to 51 percent), to be criminal (58 compared to 20 percent) and violent (30 compared to 8 percent). There are no stark differences in terms of previous breakdown, psychological problems, drug addiction or dysfunctional families between the sexes. Fifty-two percent of the teenagers in the data are boys.

Post-treatment outcomes are not good for the teenagers in the sample. At the age of 25, 73 percent of the boys in the sample had been convicted for some legal offence and 32 percent had served time in prison. The crime rate is considerably lower among girls: 27 percent had been convicted and 4 percent had been imprisoned. A clear majority – 68 percent of boys and 63 percent of girls – had not completed more schooling than the nine years of elementary education which are compulsory in Sweden.

There are some noteworthy differences between private and public HVB-homes in the data, which fit well with the model's predictions. For example, the average personnel density – an indicator of effort on quality – is 0.80 in private for-profit facilities, 1.01 in private non-profit facilities and 1.49 in public facilities. As facility size does not vary systematically with ownership (the average number of treatment places was 8.32 for public facilities and 7.64 for private facilities) the difference in personnel density is due to a larger personnel force. The average treatment cost per month was 43,497 SEK in private for-profit facilities, 48,761 SEK in private non-profit facilities and 50,645 SEK per month in public facilities.[40]

There is also a systematic difference in terms of treatment programs. Thirty-one percent of teenagers in public facilities were placed in facilities that, in addition to treatment, did evaluations of the teenagers' needs of treatment. The corresponding figures are 10 percent for private non-profit and 3 percent in private for-profit facilities. However, as teenagers who were placed in youth care with the sole purpose of having their problems evaluated have been excluded from the sample, the fact that a facility did evaluations does not imply that a certain teenager was actually placed there in order to be evaluated. The most common form of treatment program was milieu therapy in public and private non-profit facilities and to create "a family-like atmosphere" in private for-profit facilities.

The breakdown frequency was 45.3 percent in private for-profit, 29.4 percent in private non-profit facilities and 25.8 percent in public facilities.[41] However, teenagers with severe problems had higher risk of treatment breakdowns in public facilities. For example, the breakdown frequency for violent teenagers was 47.8 percent in private

[40] The SEK/Dollar exchange rate as of January 1st 1991 was 5.635.

[41] The definition of breakdown is that "a placement is ended abruptly and without planning" (Vinnerljung, Sallnäs and Kyhle-Westermark, 2001, p. 67, translated by the author). This definition thus has a clearly negative connotation.

for-profit facilities and 30.0 percent in private non-profit facilities, but 61.3 percent in public facilities.

The average duration of treatment was much longer in private for-profit facilities (21.9 months) and private non-profit facilities (18.9 months) than in public facilities (10.1 months). Part of this is reflected in a longer planned duration of treatment; which was 16.4 months in private for-profit facilities, 13.8 months in non-profit facilities and 11.0 months in public facilities. There is also a substantial difference as to the distance from the facility to the municipality center: whereas 8.1 percent of teenagers in public facilities stayed in a facility situated more than 100 kilometers away from the municipality center, this was true for 44.9 percent of the teenagers in private for-profit and 49.0 percent in private non-profit facilities.

Teenagers with severe problems are more likely to end up in private facilities: As shown in Table A1, the prevalence of teenagers with some kind of psychological disorder, drug addiction and criminal or violent behavior were all significantly higher in private facilities. Private facilities were also more likely to treat teenagers with a long treatment history. For example, almost 50 percent of teenagers in private for-profit facilities came to the facility from some other treatment program or foster care, compared to 36 percent in private non-profit facilities and 24 percent in public facilities. Regressing each teenager characteristic on an indicator variable of private ownership and a set of facility characteristics, I find that part of the selection of troublesome teenagers into private facilities is explained by distance from facility to municipality center.[42] A likely explanation for this result is that social welfare secretaries want to place teenagers with severe problems at facilities some distance from their home environment, thereby disqualifying the municipality's own facilities.[43]

To get a notion of which municipalities chose private providers, I regress a dummy variable for choice of private ownership on a set of municipality characteristics. As shown in Table B4, the propensity to choose private provision is unrelated to political majority (leftist or rightist), population size and the number of teenagers the municipality placed in HVB-homes during 1991. However, municipalities within the largest cities in Sweden – Stockholm and Gothenburg – were significantly less likely to buy youth

[42] See Table B1 to B3 in the Appendix. The set of facility characteristics includes number of treatment places and dummy variables variables for whether a facility does investigations of the teenagers' need of treatment, school at the facility and a dummy variable equal to one if the distance between the municipality center and the facility exceeds 100 km. I also include dummy variables for private nonprofit and CAB ownership.

[43] There is some support for this notion in the data. For example, 20.0 % of teenagers in public facilities situated less than 100 km from the municipality center had experienced a previous breakdown compared to 28.6 % in the corresponding private for-profit facilities. In private for-profit facilities situated more than 100 km from the municipality, 54.2 % of teenagers had experienced a previous treatment breakdown, compared to 33.3 % in the very few (15) placements in public facilities.

care from private providers, though this effect is sensitive to the exclusion of non-profit facilities from the sample. One likely explanation for this result is that Stockholm and Gothenburg are so large that they can provide an extensive set of treatment options within the municipality and county administration. The propensity to buy youth care from private providers is unrelated to the share of spending on child care and elderly care that went to private providers in 1998.[44]

6. Results

As shown above, selection of teenagers into private and public provision is not random and any empirical strategy that aims at establishing a causal effect of ownership on treatment outcomes must take this into account. Ideally, we would like to use instrumental variables to obtain exogenous variation in facility ownership, but I have not found a valid instrument.[45] I instead test if the results are due to non-random selection by subsequently adding vectors of covariates for teenager, facility and municipality characteristics to the regressions. As we will see, the results are robust to such sensitivity analysis.

I also run regressions with municipality fixed effects, thereby controlling for all variation in ψ, α, β and $\underline{q}$ that depends on municipality characteristics. It also implies that I am controlling for whether the social service is sophisticated or naive. Since most municipalities place few teenagers, controlling for municipality fixed effects adds about 100 municipality-dummies to the regressions. As the data set is not very large, this gives a substantial drop in the degrees of freedom, implying that the estimated results are less precise.

I use two dummy variables as indicators of troublesome teenagers; *Previous breakdown* and *Violent behavior*. Apart from the arguably strong a priori reasons to think that violent teenagers and those that have already experienced a breakdown are extra hard to treat, these variables have the strongest pairwise correlation with treatment breakdown among the set of teenager characteristics.

An important consideration is which facility characteristics to control for. I want to control for aspects of youth care which are correlated with – but not a consequence of – ownership. For example, I choose to include a dummy variable for whether a facility evaluates the treatment needs of teenagers as a basic control. My argument is

[44] This data is provided by Statistics Sweden. Data on the extent of service contracting in Swedish municipalities is not available prior to 1998.

[45] A valid instrument would not only have to be exogenous with respect to teenager characteristics, but also with respect to unobservable municipality and facility characteristics that affect outcomes. Among the instrument candidates I have considered are geography dummies (relevant, but hard to argue for exogeneity) and the share of other services that municipalities contract out (not relevant).

that doing an evaluation of a teenager's problems is a different service than treating these problems. The *objective* of youth care is different. Even though the data has been restricted to teenagers who also receive some form of treatment, prudence calls for including it as a control variable. In contrast, I don't control for choice of therapy as this refers not so much to objectives as to means, and is likely to be influenced by owner incentives. Similarly, I control for the geographic location of municipalities, but not the location of facilities, since strong incentives to cut cost should induce private owners (but not public managers) to seek out a locations with low costs of production.

6.1. Cost. I use personnel density (employees per treatment place) as a proxy for effort on quality.[46] There are two potential sources of bias in using personnel density as a measure of effort on quality. First, if private facilities use their personnel more efficiently – which theory suggests they will – the level of personnel density underestimates the true effort on quality in private facilities. On the other hand, private facilities have a stronger incentive to overstate their actual personnel density.[47] Since these two sources of bias contradict each other, I cannot infer the direction of the total bias.[48]

I consider the following model for the price per month municipality h pays for placing teenager i in facility j

$$
\begin{aligned}
Price_{hij} \;=\; & \alpha_{hij} + \beta_1 Personneld_j + \beta_2 Private_j + \beta_3 Private_j * Personneld_j \\
& + \beta_4 Nonprofit_j + \beta_5 CAB_j + \mathbf{X}\boldsymbol{\beta}_{\mathbf{X}} + \mathbf{Y}\boldsymbol{\beta}_{\mathbf{Y}} + Private_j * \mathbf{Y}^{*}\boldsymbol{\beta}_{\mathbf{Y}^{*}} \\
& + \mathbf{Z}\boldsymbol{\beta}_{\mathbf{Z}} + \varepsilon_j + \varepsilon_i,
\end{aligned} \tag{6.1}
$$

where $Personneld_j$ is the personnel density of facility j, $Private_j$ is a dummy variable equal to one when j is a private facility, $Nonprofit_j$ is a dummy equal to one when a private facility is owned by a non-profit organization and CAB_j is a dummy equal to one when a public facility is owned by a CAB. $\mathbf{X}$ is a vector of facility characteristics,

[46] In private facilities, personnel density is a strong predictor of treatment breakdowns, indicating that personnel density is an important aspect also of actual quality. An increase in personnel density by 1.0 predicts a reduction in breakdown frequency by 18.0 percentage units in a linear regression without control variables. The effect is strenghtened to 22.9 percentage units when I include the full set of teenager characteristics. Both coefficients are statistically significant at the 5 %-level. There is no economically or statistically significant relationship between treatment breakdowns and personnel density in public facilities.

[47] The difficulty in estimating the actual personnel density was stressed by several of the interviewed social welfare secretaries.

[48] A further potential problem with personnel density as a measure of quality is that the public owner can order the manager to employ any number of people, which seemingly contradicts the assumption that there is an upper bound on effort on quality under public provision. However, for more personnel to have a beneficial effect on quality, the manager must arguably work harder. If the manager refuses to put in such effort, the extra personnel may have no effect on quality.

including number of treatment places, treatment characteristics (school at the facility and whether the facility did evaluations of teenagers' need of treatment), and a dummy variable equal to one when distance between facility j and municipality h exceeded 100 km. Y is a vector of teenager characteristics, including sex, age, immigrant status, treatment history and problem background.[49] $\mathbf{Y^*}$ is a subset of $\mathbf{Y}$ with the teenager characteristics that denote particularly troublesome teenagers, i.e., teenagers with violent behavior and previous breakdown. $\mathbf{Z}$ is a vector of municipality characteristics, including the logarithm of municipality population in 1990, a dummy for right-wing political majority in the municipality council and six regional dummies. ε_j is an unobserved facility fixed-effect and ε_i is an individual level error term. I thus assume an additive error structure.

If my hypothesis is correct, β_2 (the difference in intercept on a price-personnel density line) should be negative and β_3 (the difference in slope) should be positive.[50] The estimate of β_2 will be biased upward if personnel density underestimates the true effort on quality in private facilities, or biased downward if personnel density overestimates effort on quality in private facilities. In addition, β_2 could be biased if expected duration of treatment differ between private and public provision. I include $Nonprofit_j$ and CAB_j to control for different intercepts between the two ownership sub-categories. The interaction effects $\mathbf{Y^*} * Private_j$ are added as theory predicts that private facilities will put more effort on troublesome teenagers. The estimate of β_2 thus only refers to non-troublesome teenagers when the interaction between ownership and troublesome teenagers is controlled for.

Table 1 gives the results for five different specifications of regression (6.1). Private firms are cheaper for low levels of quality, i.e., β_2 is negative. In the specifications without municipality fixed effects, private for-profit facilities are between 10,066 and 14,860 SEK cheaper than municipality facilities for a personnel density of "zero" and the effect is statistically significant in three out of four regressions. The difference between private-for-profit and CAB facilities ($\beta_2 - \beta_5$) is negative as expected and statistically significant in specifications three and four which include controls for facility characteristics. The interaction effect, β_3, has the expected positive sign and is statistically significant in all specifications. Depending on the exact specification, an increase in personnel density by 1.0 increases the price by between 15,505 and 20,374 SEK more in private than in public facilities. The estimate of β_2 drops to -5,022 SEK and is not statistically significant when I control for municipality fixed effect, but the

[49] See Table A1 for a complete list of the variables included in $\mathbf{Y}$.

[50] A potential problem with this specification is that we have no observation for zero personnel density. The difference in intercept, β_2, should therefore not be interpreted literally.

interaction effect remains sizable and statistically significant. There is no evidence of cost differences between for-profit and non-profit private facilities.[51]

TABLE 1. Cost per treatment month

	Price (SEK)				
Variable	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Constant	29,165***	26,819***	27,557***	40,892***	21,266**
	(8,438)	(9,560)	(7,674)	(13,073)	(12,651)
Personnel density	15,642***	17,680***	13,869***	11,148***	16,052**
	(5,883)	(6,422)	(4,704)	(4,046)	(8,742)
Private	-12,350*	-10,066	-11,490*	-14,860**	-5,022
	(8,895)	(9,841)	(7,986)	(7,734)	(12,109)
Private* pers. density	17,682***	15,505**	18,828***	20,374***	15,423**
	(6,673)	(7,366)	(5,592)	(5,817)	(9,158)
Nonprofit	-1,956	-1,617	366	1,256	8,960
	(4,147)	(4,246)	(4,008)	(4,838)	(7,587)
CAB	-1,467	-919	-466	-502	5,704
	(3,356)	(3,602)	(3,163)	(4,545)	(5,646)
Private*violence		811	571	944	2,504
		(3,505)	(3,357)	(3,929)	(4,074)
Private*prev. breakdown		6,791*	2,953	1,935	1,056
		(4,991)	(4,574)	(4,852)	(3,974)
Teenager charact.	No	Yes	Yes	Yes	Yes
Facility charact.	No	No	Yes	Yes	Yes
Municipality charact.	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
p-value* $\widehat{\beta}_2 < \widehat{\beta}_5$	.15	.24	.09	.04	.21
N	302	289	288	258	263
Number of clusters	127	125	124	114	114
R^2	.36	.38	.50	.57	.82

The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

According to the model, the fact that a teenager is troublesome may not raise prices in private facilities for high levels of quality, whereas prices rise for all levels of quality in public facilities. To control for a potential interation effect between ownership,

[51] I have tested for non-linearities in personnel density by including a quadratic term in regression (6.1). The quadratic term was never statistically significant and the adjusted R^2 decreased in three out five specifications. These results are not presented in the paper but are available from the author upon request.

troublesome teenagers and general quality, I rerun all five specifications of regression (6.1) restricting the sample to teenagers that are not violent or have experienced a previous breakdown. The results are very similar compared to those in Table 1, with the exception that the estimate of β_2 is -19,858 SEK and statistically significant in specification five with municipality fixed effects.[52]

Table A2 shows the results from additional specification checks where I use specification four with controls for teenager, facility and municipality characteristics as the base case. First, I interact personnel density with non-profit and CAB facility ownership and add these variables to the regression. Neither of these extra interaction effects are statistically significant, though the interaction between non-profit facility and personnel density is positive and quite large. Second, I add an interaction effect between private ownership and right-wing political majority in municipality h to control for political preferences for private provision. This effect is negative and statistically significant, implying that right-wing municipalities pay less for privately provided youth care. Third, I add an interaction effect between private provision and the size of the population in municipality h.[53] This effect is negative and statistically significant, indicating that large municipalities have market power. The results for β_2 and β_3 are robust to these specification tests. To check that the results for β_3 is not an artefact of selection on quality level, I rerun all five specifications of regression (6.1) restricting the sample to facilities with a personnel density of at least 0.8, thereby excluding 73 out of 155 placements in private facilities. As shown in Table B5, the interaction effect is weaker in the specifications without controls for facility characteristics, but stronger in the specifications with the full set of controls and municipality fixed effects.[54]

As a further robustness check, I calculate the expected duration of treatment for each placement conditioning on ownership and teenager characteristics and include it into regression (6.1).[55] Whereas the interaction effect (β_3) is robust to controlling for the expected duration of treatment, the intercept (β_2) drops about 25 % in specification one, three and four, and is close to zero in specification two and five. The expected duration of treatment has a negative sign in all specifications and is statistically significant in three out of five specifications. However, a closer look at the data reveals that the effect of expected duration of treatment on cost is only present for public facilities. As shown in Table B6, including an interaction effect between ownership and expected

[52] These results are not presented in the paper but are available from the author upon request.

[53] To facilitate comparison with the specifications presented above, I normalize the interaction term around the mean of municipality size.

[54] Note that the ownership dummy has no causal interpretation when we restrict the sample to facilities with a personnel density of at least 0.8.

[55] I regress duration of treatment on the full set of control variables for teenager characteristics for private and public ownership separately.

duration in the regression, I find that private facilities are cheaper for low levels of quality when the expected duration is short, but have similar cost as public facilities for a duration of about 20 months. This is in line with theory. The model predicted that cost per day should be lower for teenagers with a long expected duration of treatment under public provision, whereas under private provision prices for positive levels of quality is unaffected by expected duration of treatment if α is low and facilities earn rents.

Another potential source of bias in regression (6.1) is that stated costs in the public sector do not reflect true costs. As seen in Figure 1 and Figure 2, the correlation between cost and personnel density is much stronger for private than for public facilities. This is due to a low correlation for CAB facilities (.218) and not for municipality facilities (.789).[56] A potential explanation for this discrepancy is that some CABs sponsor their facilities, thereby weakening the link between personnel density and cost. If subsidies are positively correlated with personnel density, then β_3 will be biased upward. As a further robustness check, I therefore run regression (6.1) excluding CAB facilities. Due to the lower number of observations, these estimates (reported in Table B7) are less precise than those in Table 1. The intercept (β_2) is negative in all specifications, but not statistically significant in the two most basic specification without controls for facility and municipality characteristics. The interaction effect (β_3) is positive and statistically significant in all specifications except specification two where I only control for teenager characteristics.

FIGURE 1. Private facilities

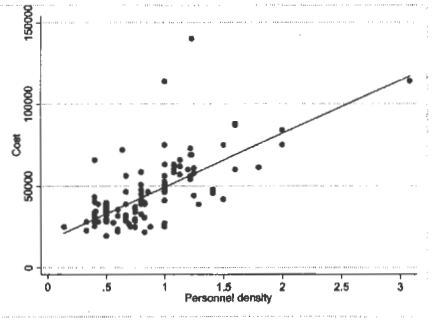
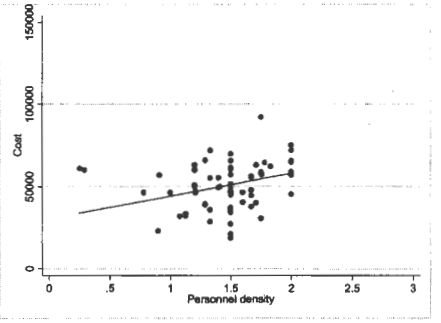


FIGURE 2. Public facilities



[56] The pairwise correlation for private for-profit facilities is .690 and .514 for non-profit facilities. The relatively strong correlation between prices and personnel density in the municipality facilities is reassuring as it indicates that prices track costs. As shown in Table B9, CAB and municipality facilities have practically identical mean values for personnel density and cost of treatment, but CAB facilities have a much lower variance in personnel density and higher variance in cost.

Though few teenagers in public facilities come from municipalities or counties that don't have a stake in the facility, a related concern is that prices in the public sector differ between "insiders" and "outsiders". To test for this, I rerun regression (6.1) including an indicator variable that takes the value one when a teenager in a CAB facility comes from a different county and an indicator variable that takes the value one when a teenager in a municipality facility comes from a different municipality. As shown in Table B8, the results only change marginally compared to the base case.

6.2. Effort on quality. I measure effort on quality by personnel density. As discussed above, we cannot be sure whether or not personnel density is a biased measure of quality in private facilities. I use the following regression for personnel density in facility j with teenager i from municipality h:

$$\begin{aligned} Personneld_{hij} = \; & \alpha_{hij} + \beta_1 Private_j + \beta_2 Nonprofit_j + \beta_3 CAB_j \\ & + \mathbf{X}\boldsymbol{\beta}_{\mathbf{X}} + \mathbf{Y}\boldsymbol{\beta}_{\mathbf{Y}} + Private_j * \mathbf{Y}^*\boldsymbol{\beta}_{\mathbf{Y}^*} + \mathbf{Z}\boldsymbol{\beta}_{\mathbf{Z}} \\ & + \varepsilon_j + \varepsilon_i, \end{aligned} \tag{6.2}$$

where $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Y}^*$ and $\mathbf{Z}$ are the same vectors of control variables as above.

As shown in Table 2, private for-profit facilities have on average .866 fewer employees per treatment place than municipality facilities when I control for for the full set of covariates, and the coefficient is robust to different specifications.[57] Including municipality fixed effects strengthens the estimated effect to 1.055 employees per treatment place. This is reassuring as one concern is that observed differences between private and public facilities depend on non-random selection of municipalities. Non-profit private facilities have a bit higher personnel density than for-profit facilities, the coefficient varies between .17 and .20 depending on the specification and is statistically significant except when I control for municipality fixed effects. There is no statistically significant difference between public facilities run by municipalities and CABs. I find no evidence of the non-monotonicity in terms of effort on quality suggested by theory: the share of private facilities does not increase as I restrict the sample to the highest levels of personnel density.[58]

There are other measures of effort on quality than personnel density. For example, in an extensive literature study of the research on youth care, Andreassen (2003)

[57] Including interaction effects, $Private_j * Y^*$, implies that I am in effect controlling for t_u. The interpretation of specification (1) is thus slightly different from (2) - (5); the first specification estimates the unconditional difference in effort on general quality, whereas the other specification estimates the difference in genereral quality, conditioning on t_u.

[58] These results are available from the author upon request.

TABLE 2. Personnel density

	Personnel density				
Variable	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Constant	1.517***	1.419***	1.328***	1.699***	1.514***
	(.169)	(.201)	(.205)	(.365)	(.340)
Private	-.720***	-.730***	-.680***	-.866***	-1.055***
	(.180)	(.192)	(.192)	(.226)	(.397)
Nonprofit	.217**	.222**	.237***	.239**	.167
	(.105)	(.097)	(.087)	(.103)	(.190)
CAB	-.013	-.002	-.011	-.122	-.307
	(.175)	(.179)	(.176)	(.207)	(.332)
Private*violence		-.073	-.088	.034	.074
		(.122)	(.112)	(.145)	(.218)
Private*prev. breakdown		.065	.130	.225	.443*
		(.113)	(.116)	(.149)	(.238)
Teenager characteristics	No	Yes	Yes	Yes	Yes
Facility characteristics	No	No	Yes	Yes	Yes
Municipality characteristics	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
p-value* $\widehat{\beta}_2 < \widehat{\beta}_5$	.00	.00	.00	.00	.00
N	328	312	311	279	284
Number of clusters	150	145	144	133	133
R^2	.34	.37	.39	.47	.66

The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

suggests that properly educated personnel is one of the most important factors for successful treatment. Unfortunately, the data set does not include information on the educational level of the personnel. Yet in a study of HVB-homes in the middle of the 1990's, Sallnäs (2000) found that half of the private for-profit facilities lacked personnel with "core competence" (defined as an education in psychology or social work), whereas this was true for 17 % of private non-profit facilities and only 5 % of public facilities.

6.3. Treatment breakdowns. The probit regression for the breakdown probability of the placement of teenager i from municipality h in facility j is

$$P\left(Break_{hij}=1 \mid x\right)=\Phi\left(\begin{array}{l}\alpha_{hij}+\beta_1 Private_j+\beta_2 Nonprofit_j+\beta_3 CAB_j \\ +\mathbf{X}\boldsymbol{\beta}_{\mathbf{X}}+\mathbf{Y}\boldsymbol{\beta}_{\mathbf{Y}}+Private_j * \mathbf{Y}^{*}\boldsymbol{\beta}_{\mathbf{Y}^{*}}+\mathbf{Z}\boldsymbol{\beta}_{\mathbf{Z}} \\ +\varepsilon_j+\varepsilon_i,\end{array}\right), \tag{6.3}$$

where $Break_{ij}$ is a dummy variable equal to one when treatment breaks down and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Y}^*$ and $\mathbf{Z}$ are the same vectors of control variables as above.[59]

As shown in Table 3, the effect of private for-profit ownership on the breakdown frequency of non-troublesome teenagers (β_1) has the expected positive sign in all specifications. Controlling for teenager, facility and municipality characteristics, a non-troublesome teenager face a 25.4 percentage units higher risk of treatment breakdown in a private for-profit facility than in a municipality facility.[60] The estimated effect is robust to the inclusion of facility, teenager and municipality characteristics, though it is not statistically significant in the regression with municipality fixed effects due to an increase in standard errors. Nonprofit private facilities have lower breakdown frequency for non-troublesome teenagers than private for-profit facilities, but the coefficient is statistically significant only in the specifications without controls and in the municipality fixed-effects regressions. The difference between private for-profit and CAB facilities ($\beta_1 - \beta_3$) is statistically significant in all specifications.

As predicted by theory, the interaction effect between private ownership and violent teenagers is negative and statistically significant. In the specification with the full set of control variables (except municipality fixed effects), the fact that a teenager is "violent" increases the risk of breakdown with 26.1 percentage units more in a public than in a private facility.[61] The corresponding figure for previous breakdown is 10.6 percentage units, but the effect is not statistically significant. As shown in Table B10, the results are very similar when I rerun regression (6.3) excluding private non-profit facilities from the sample.

The theoretical model assumed that the probability of treatment breakdown is independent of the duration of treatment. This is not an innocuous assumption. In particular, we would expect the probability of treatment breakdown to increase with the duration of treatment. If duration of treatment is longer in private facilities, the breakdown frequency may therefore underestimate quality under private provision. To correct for the possible bias arising from the private facilities' stronger incentives to prolong treatment, I calculate the expected duration of treatment conditioning on ownership and teenager characteristics and include that as a covariate in regression (6). As shown in Table B11, the results are robust to this specification test and the

[59] I use OLS in the regression with municipality fixed-effects as the probit regression is overfitted in that case.

[60] The figure refers to the average partial effect. A linear regression on the full specification (3) gives an estimate of 22.7 percent.

[61] This refers to the average partial effect. A linear regression gives an estimate of 37.2 percentage units.

effect of expected duration on the probability of treatment breakdown is small and not statistically significant.[62]

TABLE 3. Treatment breakdowns

	Treatment breakdowns (APE)			
Variable	(1) Probit	(2) Probit	(3) Probit	(4) OLS
Private	.272***	.260***	.254**	.249
	(.081)	(.099)	(.124)	(.180)
Nonprofit	-.138*	-.098	-.084	-.251*
	(.071)	(.075)	(.077)	(.131)
CAB	.108	.044	.025	-.044
	(.088)	(.092)	(.117)	(.158)
Private*violence		-.289***	-.261***	-.508***
		(.053)	(.049)	(.185)
Private*prev. break		-.073	-.106	-.064
		(.101)	.105	(.196)
Teenager characteristics	No	Yes	Yes	Yes
Facility characteristics	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	No
Municipality FE	No	No	No	Yes
p-value* $\widehat{\beta}_2 > \widehat{\beta}_5$	.00	.00	.00	.01
N	338	321	281	286
Number of clusters	158	153	134	134
Psuedo R^2/R^2	.03	.17	.23	.58

The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test, except for Nonprofit and CAB which refers to a two-sided test.

A concern regarding the interaction effect between ownership and troublesome teenagers is that whereas the proportion of teenagers that are troublesome teenagers is higher in private facilities, public facilities may treat the very worst cases. For example, "violent" teenagers in public facilities may be more violent than "violent" teenagers

[62] I get similar results if I instead include planned duration of treatment as a control variable. As the data only contains 108 observations on planned duration of treatment, the sample size is cut by approximately two thirds in these specifications. The underlying (probit) coefficients are still statistically significant and have the expected sign in specification two and three, though the average partial effect is insignificant in specification three. Planned duration of treatment has a statistically significant positive effect on treatment breakdowns in specification one, but is insignificant and even negative once control variables are included. These results are not reported in the paper but are available from the author upon request.

in private facilities. If this is the case, we would expect a higher proportion of violent teenagers in public facilities to display other problems, such as criminal behavior or drug addiction. In fact, as shown in Table B12, the proportion of violent teenagers displaying any particular problem included in the vector of teenager characteristics is higher in private facilities, suggesting that the interaction effect is actually biased toward zero.[63]

To get a more complete picture of the causes treatment breakdown, I consider data on who initiated treatment breakdowns. Private for-profit facilities initiated 29.6 % of treatment breakdowns for troublesome teenagers (teenagers with a previous breakdown, violent behavior, or both) and 36.8 % of treatment breakdowns for non-troublesome teenagers. In comparison, public facilities initiated 60 % of treatment breakdowns for troublesome teenagers and 23.5 % of treatment breakdowns for non-troublesome teenagers. The difference between private and public facilities in the proportion of breakdowns initiated by the facility is statistically significant for troublesome teenagers, but not for non-troublesome teenagers. The difference between troublesome and non-troublesome teenagers is statistically significant for public facilities, but not for private facilities. These results are consistent with the view that public facility managers are less willing to put effort on troublesome teenagers. Facing a decision between putting extra effort on a troublesome teenager or initiating a treatment breakdown, we would expect managers in public facilities to opt for the latter. However, it could also be explained by differences in manager or owner willingness to initiate a treatment breakdown when outcomes are bad, i.e., $\underline{q}$ in the theoretical model. In particular, we might worry that the breakdown frequency for troublesome teenagers is suboptimally low in private facilities. To control for differences in facility initiative, I redid the analysis excluding observations where treatment breakdowns occurred on the facility's initiative. As shown in Table B13, the interaction term between private ownership and violent teenagers is still negative and statistically significant, whereas the interaction term with previous breakdown is positive, but close to zero and not statistically significant. The direct effect of ownership is positive as before, but not statistically significant in the regressions with the full set of control variables and municipality fixed effects.[64] However, as 44.4 % of treatment breakdowns in the sample occur on the facility's

[63] Violent teenagers from Stockholm and Gothenburg seem to be equally troublesome in private and public facilities, whereas the difference is very sharp for the rest of the sample. Excluding Stockholm and Gothenburg from the sample results in weaker interaction effects. However, since violent teenagers in private facilities have much more severe problems than violent teenagers in public facilities in this subsample, we should expect the estimated interaction effect to be biased toward zero.

[64] The results are similar if I instead treat treatment breakdowns to which facilities took the initiative as "not breakdown" and include them in the sample.

initiative, excluding them from the sample implies that the size of the estimated effect goes down even if there is no systematic difference in the propensity to initiate a treatment breakdown given a certain level of quality. Moreover, there are three other reasons to believe that the high fraction of breakdowns for troublesome teenagers initiated by public facilities does not reflect concerns for the teenagers' welfare. First, given the strong detrimental effects of treatment breakdowns on teenagers well-being documented by researchers within social work, it seems unreasonable that breakdown frequencies of 50 percent or higher (which troublesome teenagers experience in public facilities) should be optimal.[65] Second, if the breakdown frequency is suboptimally low for troublesome teenagers in private facilities, the social services should be more prone to initiate treatment breakdowns for troublesome than for non-troublesome teenagers in private facilities. In fact, the proportion of treatment breakdowns initiated by the social services does not depend on teenager characteristics.[66] Third, as shown in Section 6.5 below, troublesome teenagers also have relatively better post-treatment outcomes if treated in private facilities.

Alternatively, public facility managers may shun troublesome teenagers not because they are unwilling to put down effort, but because troublesome teenagers have a negative effect on other teenagers. A social welfare maximizer may then initiate treatment breakdowns for troublesome teenagers even if the teenager in question would be better off by staying at the facility. However, as public facilities would arguably be more likely to initiate breakdowns for the most troublesome teenagers, the pool of troublesome teenagers that remains in care should then be less troublesome on average than troublesome teenagers in private facilities. As a result, treatment breakdowns should be *less* – not more – likely for troublesome teenagers in public facilities once breakdowns on the facility's initiative are removed from the data.

Finally, it is worth mentioning that the notion that public facility managers shun troublesome teenagers is not new. As noted by the Parliamentary Auditors (2002, p. 20), one reason for creating a single principal for all §12-homes in Sweden in 1994 was the shortage of treatment for criminal and violent teenagers since facilities had the

[65] Summarizing the research on treatment breakdowns, Vinnerljung et al. (2001) write (p. 21): "In all studies – Swedish and foreign – have adolescents consistently been saying that treatment breakdowns are painful experiences that leave dwelling feelings of bitterness and guilt." [translation by the author]

[66] The proportion of treatment breakdowns initiated by the social services in private for-profit facilities was 13.0 percent for teenager with a previous breakdown compared to 14.3 percent for those without a previous breakdown. The corresponding figure for violent teenagers was 16.7 percent compared to 14.6 percent for non-violent teenagers. Due to the low number of observations, these figures are imprecise estimates of the true frequencies in the population.

possibility to deny unwanted teenagers treatment places. This indicates that Swedish government did not believe in the public facility manager as a social welfare maximizer.

6.4. Duration of treatment. To test if private facilities prolong treatment periods, I run the following regression for the duration of treatment of teenager i from municipality h in facility j:

$$\begin{aligned} Duration_{hij} = \; & \alpha_{ij} + \beta_1 Private_j + \beta_2 Nonprofit_j + \beta_3 CAB_j + \beta_4 Break_i \\ & + \mathbf{X}\boldsymbol{\beta}_{\mathbf{X}} + \mathbf{Y}\boldsymbol{\beta}_{\mathbf{Y}} + Private_j * \mathbf{Y}^{*}\boldsymbol{\beta}_{\mathbf{Y}^{*}} + \mathbf{Z}\boldsymbol{\beta}_{\mathbf{Z}} + \varepsilon_j + \varepsilon_i, \end{aligned} \tag{6.4}$$

where $Break_i$ is the breakdown-dummy from above and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Y}^{*}$ and $\mathbf{Z}$ are the same vectors of control variables as above. I control for treatment breakdowns as I want to identify the effect of ownership on duration of treatment that goes through information on treatment progress, not the effect that goes through quality.

As shown in Table 4, private for-profit ownership increases duration of treatment by between 13.0 and 14.9 months compared to municipality facilities depending on the specification and the effect is statistically significant. Non-profit private facilities have 4.40 months shorter duration of treatment than for-profit private facilities in the full regression and the difference is statistically significant. These results are robust to specification checks with exponential and Weibull duration models.[67]

The results are robust to a number of further tests, reported in Table B14. To better control for effort on quality, I add personnel density and the full set of control variables for treatment programs. The estimated partial effect of private for-profit ownership is then 12.6 months when I control for teenager, facility and municipality characteristics and 12.0 months in the regression with municipality fixed effects, and the coefficient is statistically significant in both cases.[68] Another potential cause of concern is that municipalities which desire a long period of treatment also have a low valuation of quality, i.e., that ψ and β are negatively correlated. If so, private provision may be more appealing due to its lower per-day cost and the longer duration of treatment would not be due to lack of information on treatment progress. Yet including cost per day as an additional control variable in the regression of Table 4 actually strengthens the effect of private provision on duration of treatment.[69]

[67] These results are available from the author upon request.

[68] Since private facilities may choose treatment programs that better enable them to prolong the duration of treatment, including them may imply overcontrolling.

[69] Note that cost per day of treatment is endogenous in this regression since cost is decreasing in expected duration of treatment.

TABLE 4. Duration of treatment

	Duration (months)			
Variable	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Constant	9.042***	10.385***	13.542	14.319***
	(2.099)	(2.135)	(10.780)	(5.349)
Private	13.015***	14.879***	14.223***	13.021*
	(2.660)	(2.613)	(3.266)	(6.590)
Nonprofit	-3.194	-4.294	-4.400	-.759
	(3.123)	(2.928)	(3.053)	(3.883)
CAB	1.511	2.122	1.603	-.139
	(2.453)	(2.487)	(2.669)	(5.586)
Breakdown		-7.166***	-7.502***	-9.004***
		(1.685)	(1.790)	(2.450)
Private*violence			-1.549	-1.856
			(3.257)	(5.099)
Private*prev. break			-2.921	.317
			(3.334)	(5.386)
Teenager charact.	No	No	Yes	Yes
Facility charact.	No	No	Yes	Yes
Municipality charact.	No	No	Yes	No
Municipality FE	No	No	No	Yes
p-value* $\widehat{\beta}_2 - \widehat{\beta}_5$	.00	.00	.00	.00
N	336	336	279	284
Number of clusters	157	157	133	133
R^2	.15	.21	.40	.68

The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

The model gave two different reasons for why private provision may increase the duration of treatment. First, lack of accurate information on treatment progress may induce sophisticated social services to plan for a longer duration of treatment. Second, naive social services may believe that the facility owner is actually telling the truth regarding treatment progress. The mean values considered in Section 5 supported both of these views – planned duration of treatment was about five months longer in private facilities, but duration of treatment also exceeded plan by an additional five months under private provision. To separate the two effects, I include planned duration of treatment in each of the three first specifications in Table 4. As shown in Table B15, the effect of ownership is reduced to about eight months in the specifications without controls and is close to zero and statistically insignificant in the specification

with controls for teenager, facility and municipality characteristics. However, I lose about two-thirds of the sample by including planned duration of treatment, and the discrepancy with the results for the full sample is largely driven by sample selection. I also run the duration regressions with imputed values for planned duration of treatment with results very similar to those in Table 4.[70]. In sum, the effect of private ownership on the duration of treatment seems to be mostly due to naive social welfare secretaries.

The longer duration of treatment in private facilities implies that only considering cost per month underestimates the total cost difference between private and public provision of youth care. In my sample, the average total cost of treatment was 1,045,976 SEK for private-for-profit facilities and 886,409 SEK for private non-profit facilities compared to 501,492 SEK for CAB facilities and 425,282 SEK in municipality facilities.[71]

6.5. Post-treatment outcomes. One way to evaluate post-treatment outcomes is to consider whether a teenager continued treatment in some other facility after the placement had ended. As shown in Table A3, teenagers in private facilities are less likely to continue treatment than teenagers in public facilities. Yet this effect is not robust when I control for facility characteristics. In particular, the difference in terms of continued treatment is explained by the larger share of facilities that do evaluations of teenagers' treatment needs.

As an alternative measure of the effect of treatment, I consider data on economic self-reliance, educational attainment, mental health and criminal record at the age of 25. Focusing on adult outcomes introduces another difficulty in the analysis. Whereas the previous outcome measures all pertain to a particular *placement*, adult outcomes focuses on an *individual*, with a history of treatment which may go beyond a single placement. For example, some teenagers were replaced to another facility during 1991. In addition, teenagers have different treatment histories prior to 1991 and some continue treatment after the last placement we observe in the data. When a teenager experienced more than one placement in 1991, I consider the placement with the longest duration.[72] I exclude four teenagers from the sample where the difference between the longest and second longest placement was one month or less. To control for treatment history, I add

[70] I use the *regmsng* command in STATA with *nmiss(1)*. The estimated effect of private ownership on duration of treatment was 11.97 months i specification (1), 13.82 in specification (2) and 14.26 months in specification (3).

[71] About 10 % of the difference in total cost between private and public provision is explained by the fact that placements in public facilities last shorter time the more expensive the treatment, whereas there is no relationship between cost and duration of treatment in private facilities.

[72] Only considering the longest placement implies that we are in effect restricting the range of $\underline{q}$. This in turn implies that the samples of private and public placements may not be fully representative.

an indicator variable equal to one when a teenager continued treatment in some other facility to the vector of teenager characteristics. In addition, I include an indicator variable that takes the value one if a teenager had multiple placements during 1991.

The results on post-treatment outcomes are presented in Table A3 to A5. There are no statistically significant differences between teenagers at private and public facilities in terms of reliance on social assistance, educational attainment or mental health problems between the age of 20 and 24, defined as hospital care for suicide attempt, drug addiction or psychiatric illness. However, violent teenagers are relatively less likely to be convicted for a crime between the age of 20 and 24 if treated in a private facility and the effect is statistically significant in all specifications. The interaction effect between previous breakdown and private ownership is also negative, but not statistically significant in the specification with municipality fixed effects. Moreover, though the likelihood that non-troublesome teenagers are imprisoned between the age of 20 and 24 does not depend on facility ownership, violent teenagers are relatively less likely to be imprisoned if treated in a private facility and the effect is statistically significant. The interaction effect between private ownership and previous breakdown has a positive sign, but is not statistically significant.

As a robustness check, I restrict the sample to teenagers that only experienced one placement during 1991. This restriction solves the problem of which placement to refer to each teenager. However, as the risk of treatment breakdown depends on quality, restricting the sample to teenagers who did not switch facility implies that we are in effect selecting a high-quality sample.[73] The restriction is slightly more binding for private facilities where 30.2 percent of placements were teenagers that had multiple placements compared to 28.0 percent in public facilities. As shown in Table B16 to B18, the results are similar compared to the unrestricted sample.

The results for post-treatment outcomes indicate that public provision is superior for non-troublesome teenagers since outcomes are similar but total costs substantially lower.[74] The cost-benefit analysis for troublesome teenagers is less clear-cut. Though we have seen that private facilities are better at treating troublesome teenagers relative to non-troublesome teenagers, a cost-benefit analysis should focus on the absolute effect, which depends both on general quality and effort on troublesome teenagers. To get an estimate of the absolute difference in quality I rerun the post-treatment regressions

[73] The breakdown frequency was 27.6 percent for teenagers that experienced one placement compared to 43.3 percent for teenagers with more than one placement.

[74] One important caveat to this assumption is that, as discussed in Section 2 and 6.1, costs may be underestimated for public providers. However, for this to invalidate the conclusion that public providers are superior for non-troublesome teenagers, stated cost should only amount to 50 % of true costs, which seems unlikely.

for convictions and imprisonment restricting the sample to troublesome teenagers. As shown in Table B19, troublesome teenagers are less likely to be convicted if treated in a private facility. In the specification without control variables, troublesome teenagers are 34.0 percentage units more likely to be convicted if treated in municipality facility compared to a private for-profit facility, and 12.0 percentage units more likely if treated in a CAB facility, and both these effects are statistically significant. However, the size and significance of the estimated effect of private for-profit ownership on convictions is not stable across specifications.[75] Troublesome teenagers in private for-profit facilities also have a lower risk of imprisonment compared to all other facility management types, but the size and significance of the effect is not stable across specifications in this case either. Given the uncertainty of these estimates, it is hard to quantify the benefits of sending troublesome teenagers to private facilities.

6.6. The social services' effort. I use an indicator variable for whether there existed a plan for the duration of treatment as a proxy for the social welfare secretary's effort. Researchers in youth care have argued that teenagers in youth care benefit directly from a planned duration of treatment (see Sallnäs, 2000, p. 185), but it could also be presumed that social welfare secretaries who plan the duration of treatment also put more effort on other aspects of treatment.

The proportion of placements that had a plan for the duration of treatment was similar in private (30.8 percent) and public (29.6 percent) facilities. Placements with a planned duration of treatment do not differ systematically regarding the mean and variance of the duration of treatment (indicating that α is similar), or other facility and teenager characteristics. However, the largest cities in Sweden (Stockholm, Gothenburg and Malmö) are more likely to plan the duration of treatment, and the set of municipality dummies is a stronger predictor of planned duration than the set of facility dummies.[76] This indicates that it is the social welfare secretaries that decide whether the duration is planned in advance, and not the facility personnel.

Table 5 reports the results from regressing treatment breakdowns and criminal behavior on a dummy variable for planned duration of treatment. Planned duration has a negative sign in all regressions, indicating that outcomes are better when the social services exert more effort. For private facilities, planned duration is statistically

[75] Due to the limited number of degress of freedom, I include controls for teenager and facility/municipality characteristics in separate regressions.

[76] I get an adjusted R^2 of 0.01 when I regress a dummy for planned duration of treatment on the full set of facility dummies and an adjusted R^2 of 0.004 when I regress this variable on the full set of teenager characteristics. In comparison, the set of municipality dummies has more explanatory power with an adjusted R^2 of 0.12.

significant for treatment breakdowns and convictions, and the effect is strengthened when I control for teenager, facility and municipality characteristics. The effect on imprisonment is not statistically significant. Planned duration improves outcomes also in public facilities, but the effect is not as strong as for private facilities, suggesting that social services that plan the duration of treatment also sign more complete contracts with private providers.

Including price per day of treatment in the regressions reported in Table 5 has no substantial effect beyond sample selection on the estimates effect of planned duration of treatment, indicating that social welfare secretaries that plan the duration of treatment do not have a higher willingness to pay for quality.[77]

TABLE 5. Treatment outcomes divided on ownership

Private facilities

	Treatment breakdown		Convicted 20-24		Imprisoned 20-24	
Planned duration	−0.181**	−0.219**	−0.188**	−0.376**	−0.085	−0.007
	(0.079)	(0.130)	(0.087)	(0.219)	(0.067)	(0.013)
Non-profit	−0.137	−0.084	0.038	0.120	0.096	0.058
	(0.089)	(0.116)	(0.105)	(0.173)	(0.076)	(0.070)
Teenager ch.	No	Yes	No	Yes	No	Yes
Facility charact.	No	Yes	No	Yes	No	Yes
Municipality ch.	No	Yes	No	Yes	No	Yes
N	157	121	129	89	129	86

Public facilities

	Treatment breakdown		Convicted 20-24		Imprisoned 20-24	
Planned duration	−0.019	−0.177***	−0.123*	−0.223*	−0.014	too
	(0.062)	(0.055)	(0.077)	(0.139)	(0.066)	few
CAB	0.091	−0.102	−0.011	−0.110	0.077	df
	(0.071)	(0.152)	(0.090)	(0.246)	(0.067)	
Teenager ch.	No	Yes	No	Yes	No	Yes
Facility charact.	No	Yes	No	Yes	No	Yes
Municipality ch.	No	Yes	No	Yes	No	Yes
N	181	154	131	116	131	

All estimates regards the APE from Probit regressions. Robust standard errors with cluster at the facility level. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

To see if the results for private ownership displayed above depend on the social welfare secretary's engagement, I split the sample depending on whether treatment

[77] These results are available from the author upon request.

was planned or not and rerun regression (6.1) to (6.4), controlling for teenager characteristics. The results remain qualitatively similar.[78]

7. Concluding Remarks

This paper shows that there are important differences between private and public youth care facilities. Private facilities have lower per-day cost of treatment for low levels of quality, but higher per-day cost for high levels of quality. The general level of quality is lower for private facilities, but public facilities have relatively lower quality for troublesome teenagers. Finally, private provision prolongs treatment periods, thereby increasing total costs.

Though the details of these findings are specific to residential youth care, the underlying mechanisms are probably important also in other cases of services contracting. For instance, physicians know the treatment needs of individual patients much better than politicians and bureaucrats. Based on the analysis in this paper, we would then expect private hospitals to overstate the need for treatments that they find profitable. Moreover, some patients in hospital care are more demanding than others, implying that privatization may have different effects for different groups of patients. Similar issues are likely to arise in, for example, schooling, prisons, nursing homes for elderly people, and childcare. Of course, there are also important differences between these services and residential youth care. For example, a private school may refuse to teach children from disadvantaged backgrounds if it harms their ability to attract more priviliged students.

[78] These results are available from the author upon request.

Appendix A

A.1. Owner's effort choice. The full solution to the owner's maximization problem (2) gives

$$\left(t_q^o, t_c^o\right) = \begin{cases} t_q = 0 \text{ and } t_c = \underline{t} \text{ s.t. } S'(\underline{t}) = C'(\underline{t}) & \text{if } 0 \leq \widetilde{\alpha} \leq S'(\underline{t}); \\ \{(t_q, t_c) | \ \widetilde{\alpha} = S'(t_c) = C'(t)\} & \text{if } S'(\underline{t}) < \widetilde{\alpha} < S'(0); \\ t_q \text{ s.t. } \widetilde{\alpha} = C'(t) \text{ and } t_c = 0 & \text{if } \widetilde{\alpha} \geq S'(0) \end{cases}$$

Since $S(t_c)$ is increasing and strictly concave, there exists an inverse function to $\widetilde{\alpha} = S'(t_c)$ for $0 < t_c < \underline{t}$. As $S(t_c)$ is twice continuously differentiable, the inverse function $t_c^o(\widetilde{\alpha})$ is continuously decreasing in $\widetilde{\alpha}$ for $S'(\underline{t}) < \widetilde{\alpha} < S'(0)$. Since $C(t)$ is strictly convex and increasing for $t > t^*$, $\widetilde{\alpha} = C'(t)$ is invertible for $\underline{t} < t < \overline{t}$ where $\overline{t}$ is such that $\lim_{t \to \overline{t}} C'(t) = \infty$. As $C(t)$ is twice continuously differentiable, the inverse function $t^o(\widetilde{\alpha})$ is continuously increasing in $\widetilde{\alpha}$ for $\widetilde{\alpha} > S'(\underline{t}) = C'(\underline{t})$ with $\lim_{\alpha \to \infty} t^o(\alpha) = \overline{t}$. The owner's effort on quality is given by $t_q^o(\widetilde{\alpha}) = t^o(\widetilde{\alpha}) - t_c^o(\widetilde{\alpha})$. It follows from above that $t_q^o(\widetilde{\alpha}) = 0$ for $0 \leq \widetilde{\alpha} \leq S'(\underline{t})$ and that $t_q^o(\widetilde{\alpha})$ is continuously increasing in $\widetilde{\alpha}$ for all $\widetilde{\alpha} > S'(\underline{t}) = C'(\underline{t})$ with $\lim_{\alpha \to \infty} t_q^o(\widetilde{\alpha}) = \overline{t}$. In the interior solution, we thus get t_q^o as a continuous function of $\widetilde{\alpha}$ with first derivative

$$\frac{\partial t_q(\widetilde{\alpha})}{\partial \widetilde{\alpha}} = \frac{1}{C''(t)} - \frac{1}{S''(t)} > 0$$

and second derivative

$$\frac{\partial t_q(\widetilde{\alpha})}{\partial \widetilde{\alpha}} = -\frac{C'''(t(\alpha p))}{[C''(t(\alpha p))]^3} + \frac{S'''(t(\alpha p))}{[S''(t_c(\alpha p))]^3} < 0$$

as I have assumed that $C'''(t(\widetilde{\alpha})) > 0$ and $S'''(t(\widetilde{\alpha})) > 0$.

The full solution to (2) also gives

$$t_v^o = \begin{cases} t_v = 0 & \text{if } \widetilde{\alpha} \leq D'(0) \text{ or } v = 0 \\ t_v \text{ s.t. } \widetilde{\alpha} = D'(t_v) & \text{if } D'(0) < \widetilde{\alpha} < D'(v) \text{ and } v > 0 \\ t_v = v & \text{if } \widetilde{\alpha} \geq D'(v). \end{cases}$$

Since $D(t_v)$ is increasing and strictly concave, there exists an inverse function to $\widetilde{\alpha} = D'(t_v)$ for $0 < t_v < v$. As $D(t_v)$ is twice continuously differentiable, the inverse function $t_v(\widetilde{\alpha}, v)$ is continuously decreasing in $\widetilde{\alpha}$ for $D'(0) < \widetilde{\alpha} < D'(v)$. In the interior solution, we thus get t_v as a continuous function of $\widetilde{\alpha}$ with derivative

$$\frac{\partial t_v(\widetilde{\alpha}, v)}{\partial \widetilde{\alpha}} = \frac{\widetilde{\alpha}}{D''(t_v)} > 0.$$

A.2. The social service's problem under public provision. Under public provision, the social service expects to receive accurate information on treatment progress regardless of whether he is naive or sophisticated. Since the social service has a symmetric prior distribution on D and is risk neutral, the maximand is

$$E\left[\left(B\left(t_q\left(\widetilde{\alpha}\right),\psi\right)-p\right)T^*\right].$$

Under public provision, the social service can just set t_q and t_v to any value such that $t_q + t_c = t^*$. As $w^m = 0$ and

$$p_m\left(t_q\right) = \frac{c - S\left(t^* - t_q\right)}{E\left[T^*\right]},$$

the maximization problem with respect to t_q for $v = 0$ can be written as

$$\max_{t_q} \quad \left(B\left(t_q,\psi\right) + \left(S\left(t^* - t_q\right) - c\right) / \left(\alpha t_q - \alpha \underline{q} + \beta\right)\right)\left(\alpha t_q - \alpha \underline{q} + \beta\right),$$

which can be reformulated as

$$\max_{t_q} \quad B\left(t_q,\psi\right)\left(\alpha t_q - \alpha \underline{q} + \beta\right) + S\left(t^* - t_q\right) - c.$$

The first derivative is

$$B_{t_q}\left(t_q,\psi\right)\left(\alpha t_q - \alpha \underline{q} + \beta\right) + B\left(t_q,\psi\right)\alpha - S'\left(t^* - t_q\right).$$

As $\left(\alpha t_q - \alpha \underline{q} + \beta\right)$ is strictly above zero, $B\left(t_q,\psi\right)\alpha$ is non-negative, $S'\left(0\right)$ is finite and $B_{t_q}\left(t_q,\psi\right)$ is continuously increasing in ψ with $\lim_{\psi\to\infty} B_{t_q}\left(t_q,\psi\right) = \infty$ for all $t_q > 0$, there exists a finite threshold $\widehat{\psi}$ such that the first derivative is positive for all $t_q > 0$ when $\psi > \widehat{\psi}$. Consequently, the social service will set $t_q = t^*$ and $t_c = 0$ if quality is sufficiently important.

The second derivative is

$$\underbrace{B_{t_q t_q}\left(t_q,\psi\right)\left(\alpha t_q - \alpha \underline{q} + \beta\right)}_{-} + \underbrace{2\alpha B_{t_q}\left(t_q,\psi\right)}_{+} + \underbrace{S''\left(t^* - t_q\right)}_{-},$$

which might not be negative. Hence, there can be more than one interior solution to the social services problem.

A.3. The social service's problem under private provision: Naive Social Service. Since the social service has a symmetric prior distribution on D, believes that he will receive correct information on treatment progress, and is risk neutral, the maximand is

$$\begin{aligned} & E\left[\left(B\left(t_q\left(\widetilde{\alpha}\right),\psi\right)-p\right)T^*\right] \\ = \; & \left(B\left(t_q\left(\widetilde{\alpha}\right),\psi\right)-p\right)\left(\alpha\left[t_q\left(\widetilde{\alpha}\right) + \min\left\{t_v - v, 0\right\} - \underline{q}\right] + \beta\right) \end{aligned}$$

and so the maximization problem for $v = 0$ and $\alpha > 0$ can be written as

$$\begin{array}{ll} \max\limits_{p} & \left(B\left(t_q\left(\alpha p\right), \psi\right) - p\right)\left(\alpha t_q\left(\widetilde{\alpha}\right) - \alpha \underline{q} + \beta\right) \\ \text{s.t.} & p \geq p^*. \end{array}$$

If the participation constraint does not bind, we get the first derivative in the interior solution

$$\left[B_{t_q}\left(t_q\left(\widetilde{\alpha}\right), \psi\right)\left[\frac{1}{C''\left(t\left(\alpha p\right)\right)} - \frac{1}{S''\left(t\left(\alpha p\right)\right)}\right]\alpha - 1\right]\left(\alpha t_q\left(\alpha p\right) - \alpha \underline{q} + \beta\right)$$
$$+ \left(B\left(t_q\left(\alpha p\right), \psi\right) - p\right)\left[\frac{1}{C''\left(t\left(\alpha p\right)\right)} - \frac{1}{S''\left(t\left(\alpha p\right)\right)}\right]\alpha^2.$$

The second derivative in the interior solution (with sign of each term) is

$$\underbrace{B_{t_q t_q}\left(t_q\left(\alpha p\right), \psi\right)\alpha^2}_{-}\underbrace{\left[\frac{1}{C''\left(t\left(\alpha p\right)\right)} - \frac{1}{S''\left(t_c\left(\alpha p\right)\right)}\right]^2}_{+}\underbrace{\left(\alpha t_q\left(\alpha p\right) - \alpha \underline{q} + \beta\right)}_{+}$$
$$\underbrace{-B_{t_q}\left(t_q\left(\alpha p\right), \psi\right)\alpha^2}_{+}\underbrace{\left[\frac{C'''\left(t\left(\alpha p\right)\right)}{\left[C''\left(t\left(\alpha p\right)\right)\right]^3} - \frac{S'''\left(t\left(\alpha p\right)\right)}{\left[S''\left(t_c\left(\alpha p\right)\right)\right]^3}\right]}_{+}\underbrace{\left(\alpha t_q\left(\alpha p\right) - \alpha \underline{q} + \beta\right)}_{+}$$
$$\underbrace{+2B_{t_q}\left(t_q\left(\alpha p\right), \psi\right)\alpha^3}_{+}\underbrace{\left[\frac{1}{C''\left(t\left(\alpha p\right)\right)} - \frac{1}{S''\left(t_c\left(\alpha p\right)\right)}\right]^2}_{+} \underbrace{- 2\,\alpha^2\left[\frac{1}{C''\left(t\left(\alpha p\right)\right)} - \frac{1}{S''\left(t_c\left(\alpha p\right)\right)}\right]}_{+}$$
$$\underbrace{+\alpha^3\left[p - B\left(t_q\left(\alpha p\right), \psi\right)\right]}_{-}\underbrace{\left[\frac{C'''\left(t\left(\alpha p\right)\right)}{\left[C''\left(t\left(\alpha p\right)\right)\right]^3} - \frac{S'''\left(t_c\left(\alpha p\right)\right)}{\left[S''\left(t_c\left(\alpha p\right)\right)\right]^3}\right]}_{+}.$$

The reason we cannot unambiguously sign the second derivative is that the marginal benefit of higher quality increases in the duration of treatment, and prices affect the duration of treatment positively. However, the optimal p is always finite. To see this, note that for all $\alpha > 0$, there exists a finite p such that $\widetilde{\alpha} > S'(0)$ for all prices above this level. This implies that we get the corner solution where $t_c = 0$. The first derivative of the social service's problem w.r.t. p is then

$$\left[B_{t_q}\left(t_q\left(\alpha p\right), \psi\right)\frac{1}{C''\left(t\left(\alpha p\right)\right)}\alpha - 1\right]\left(\alpha t_q\left(\alpha p\right) - \alpha \underline{q} + \beta\right)$$
$$+ \left(B\left(t_q\left(\alpha p\right), \psi\right) - p\right)\frac{\alpha^2}{C''\left(t\left(\alpha p\right)\right)}.$$

As $C'''(t) > 0$, with $\lim_{t \to \bar{t}} C''(t) = \infty$ and $\lim_{\widetilde{\alpha} \to \infty}\left(t\left(\widetilde{\alpha}\right)\right) = \bar{t}$, we get that $C''(t)$ is continuously increasing in $\widetilde{\alpha}$ with $\lim_{\widetilde{\alpha} \to \infty} C''\left(t\left(\widetilde{\alpha}\right)\right) = \infty$. It follows from strict

concavity of $B(t_q, \psi)$ and $t_q(\widetilde{\alpha})$ that $B_{t_q}(t_q(\alpha p), \psi)$ is continuously falling in p. As a result, $B_{t_q}(t_q(\alpha p), \psi) < C''(t(\alpha p))$ for some finite p. Moreover, as $B(t_q(\alpha p), \psi)$ has a finite upper bound, there exists some finite p such that $B(t_q(\alpha p), \psi) < p$. Hence, there exists a finite threshold, p', such that the first derivative is negative for all $p > p'$.

For $v > 0$ and $t_v^o(\widetilde{\alpha}) < v$, we get

$$\begin{array}{ll} \max\limits_{p} & \left(B(t_q(\widetilde{\alpha}), \psi) - p\right)\left(\alpha t_q(\widetilde{\alpha}) + \alpha t_v(\widetilde{\alpha}) - \alpha v - \alpha \underline{q} + \beta\right) \\ \text{s.t.} & p \geq p^*. \end{array}$$

If the participation constraint does not bind, we get the first-order derivative

$$\begin{aligned} &\left[B_{t_q}(t_q(\widetilde{\alpha}), \psi)\frac{\partial t_q(\widetilde{\alpha})}{\partial p} - 1\right]\left(\alpha t_q(\widetilde{\alpha}) + \alpha t_v(\widetilde{\alpha}) - \alpha v - \alpha \underline{q} + \beta\right) \\ &+ \left(B(t_q(\widetilde{\alpha}), \psi) - p\right)\left(\alpha \frac{\partial t_q(\widetilde{\alpha})}{\partial \widetilde{\alpha}}\frac{\partial \widetilde{\alpha}}{\partial p} + \alpha \frac{\partial t_v(\widetilde{\alpha})}{\partial \widetilde{\alpha}}\frac{\partial \widetilde{\alpha}}{\partial p}\right). \end{aligned}$$

Hence, compared to the case of non-troublesome teenagers, the benefit of raising prices is stronger when teenagers are troublesome. As I have already shown that we cannot sign the second derivative (unless we impose additional assumptions) in the case of non-troublesome teenagers, I do not derive the second derivative also in this case.

A.4. The social service's problem under private provision: Sophisticated Social Service. If the social service knows that it will not receive accurate information on treatment progress, decisions on p and A are made simultaneously. The maximand for p is then different depending on the decision to shorten or prolong treatment. In case treatment is shortened, the maximand is

$$\left(B(t_q(\alpha p), \psi) - p\right)\left(E[T^*] - \gamma\right) \tag{A.1}$$

where

$$E[T^*] = \alpha t_q(\widetilde{\alpha}) - \alpha \underline{q} + \beta + \alpha \min\left\{t_v(\widetilde{\alpha}) - v, 0\right\}.$$

If treatment is neither shortened, nor prolonged, the maximand is

$$\begin{aligned} &\theta_{-1}\left(B(t_q(\alpha p), \psi) - p\right)\left(E[T^*] - \gamma\right) - \theta_{-1} p\gamma \\ &+ (\theta_0 + \theta_1)\left(B(t_q(\alpha p), \psi) - p\right)E[T^*]. \end{aligned} \tag{A.2}$$

where θ_{-1} denotes the probability that $D = -1$, θ_0 the probability that $D = 0$ and θ_1 the probability that $D = 1$. Finally, if treatment is prolonged, the maximand is

$$\begin{aligned} &\theta_{-1}\left(B(t_q(\alpha p), \psi) - p\right)\left(E[T^*] - \gamma\right) - 2\theta_{-1} p\gamma \\ &+\theta_0\left(B(t_q(\alpha p), \psi) - p\right)E[T^*] - \theta_0 p\gamma \\ &+\theta_1\left(B(t_q(\alpha p), \psi) - p\right)\left(E[T^*] + \gamma\right). \end{aligned} \tag{A.3}$$

Since $\theta_{-1} = \theta_1$ by assumption, (A.3) can be reformulated as

$$(B(t_q(\alpha p), \psi) - p) E[T^*] - p\gamma.$$

Further, (A.1) can be rewritten as

$$(B(t_q(\alpha p), \psi) - p) E[T^*] - (B(t_q(\alpha p), \psi) - p)\gamma,$$

and (A.2) as

$$(B(t_q(\alpha p), \psi) - p) E[T^*] - \theta_{-1} B(t_q(\alpha p), \psi)\gamma.$$

These maximands reveal that the (perceived) surplus under private provision is always lower for the sophisticated than for the naive social service. Removing the common term (which is the same as the maximand without uncertainty), the remaining terms of these maximands are

$$\begin{aligned} -(B(t_q(\alpha p), \psi) - p)\gamma \text{ if } A &= -1, \\ -\theta_{-1} B(t_q(\alpha p), \psi)\gamma \text{ if } A &= 0, \\ -p\gamma \text{ if } A &= 1. \end{aligned}$$

The basic properties of this problem (regarding the existence of solutions), is the same as for the naive social service's problem.

A.5. Proof of Proposition 1. From the solution to (3.1), we know that the maximum level of quality under public provision is $t_q^m = t^*$. Since $\overline{t} > t^*$, there exists some $S'(\underline{t}) < \widehat{\alpha} < \infty$ by continuity of $t_q^o(\widetilde{\alpha})$ such that $t_q^o(\widehat{\alpha}) = t^*$ and $t_q^o(\widetilde{\alpha}) > t^*$ for all $\widetilde{\alpha} > \widehat{\alpha}$. As $\widetilde{\alpha} = \alpha p$, there exists some p for every $\alpha > 0$ such that $\widetilde{\alpha} > \widehat{\alpha}$.

A.6. Proof of Proposition 2. From the solution to (3.1), we know that $t_v^m = 0$. From the solution to (3.2), we know that the owner sets $t_v^o > 0$ if $\widetilde{\alpha} > D'(0)$.

A.7. Proof of Proposition 3. For each level of quality and corresponding duration of treatment, there is a reservation price, denoted by $p^*(t_q)$, where the participation constraint binds, i.e., such that $u^o = 0$. When teenagers are non-troublesome and $A = 0$, this is

$$p_o^*(t_q) = \frac{c + C(t(\widetilde{\alpha}(t_q))) - S(t_c(\widetilde{\alpha}(t_q)))}{\alpha(t_q - \underline{q}) + \beta}.$$

Cost under public provision is given by

$$p_m(t_q) = \frac{c - S(t^* - t_q)}{\alpha(t_q - \underline{q}) + \beta}.$$

As $\beta = \beta' - \alpha\left(E[t_q] - \underline{q}\right)$, both p_o^* and p_m are independent of α and $\underline{q}$. For fixed c and β' we then get p_o^* and p_m as a continuously increasing functions of t_q. As $c > [C(\underline{t}) - S(\underline{t})]$, $p_o^*(t_q)$ is always strictly above zero. Since $[C(t(t_q)) - S(t_c(t_q))] < S(t^* - t_q)$, p_o^* is lower than p_m for all feasible levels of quality.

Now consider the results in the proposition. First, we know from the owner's maximization problem that $t_q^o(\widetilde{\alpha}) = 0$ for all $\widetilde{\alpha} \leq S'(\underline{t})$. Hence, $t_q^o(\alpha p_o^*(0)) = 0$ for all $\alpha \leq \widehat{\alpha}_1$ where $\widehat{\alpha}_1 = S'(\underline{t})/p_o^*(0)$. When $\alpha > \widehat{\alpha}_1$, $t_q^o(\alpha p_o^*(0)) > 0$, and so $t_q = 0$ is not feasible under private provision.

Second, for all positive levels of quality, there exists a cutoff $\widehat{\alpha}_2$ such that $t_q(\widehat{\alpha}_2 p_o^*(t_q)) = t_q$. For $\alpha > \widehat{\alpha}_2$, $t_q(\alpha p_o^*(t_q)) > t_q$, which is not feasible. As each positive level of quality corresponds to a unique $\widetilde{\alpha}$, $p_o = \widetilde{\alpha}/\alpha$ approach infinity continuously as α goes to zero. Hence, for each level of quality, there exists another unique cutoff $\overline{\alpha}_2$ such that $p_o > p_m$ for all $0 < \alpha < \overline{\alpha}_2$, but $p_o(t_q) < p_m(t_q)$ for $\overline{\alpha}_2 < \alpha \leq \widehat{\alpha}_2$.

A.7.1. *Note on information and troublesome teenagers.* As effort on general quality is just a function of αp under private provision, the price needed to extract any given positive level of quality from a private firm for a given α is not affected by the extent of a teenager's problems (v) or whether treatment is prolonged or shortened. However, as the reservation price is the lowest price that is feasible for a given level of quality, it does affect the set of α (and hence also the price levels) for which a given level of quality is feasible.

First, consider troublesome teenagers. As $t_v^m = 0$, we get

$$p_m(t_q) = \frac{c - S(t^* - t_q)}{\beta' - \alpha v},$$

implying that $p_m(t_q | v > 0) > p_m(t_q | v = 0)$.

In case of private provision, we get

$$p_o^*(t_q) = \frac{c + C(t(\widetilde{\alpha}(t_q))) - S(t_c(\widetilde{\alpha}(t_q))) + D(t_v(\widetilde{\alpha}(t_q)))}{\beta' - \alpha[t_v(\widetilde{\alpha}(t_q)) - v]}.$$

It is clear that $p_o^*(t_q | v > 0) > p_o^*(t_q | v = 0)$ for all levels of quality. However, we still have that $p_o^*(t_q) < p_m(t_q)$ for all levels of quality. To see this, note that the owner can do no worse than setting $t_v = 0$, in which case $p_o^*(t_q) < p_m(t_q)$ by the same argument as above. That $p_o^*(t_q)$ is increasing in v implies that $p_o^*(0)$ increases, and that the cutoffs $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ fall. As $p_m(t_q)$ increases, the cutoff $\overline{\alpha}_2$ also falls.

Second, consider information on treatment progress. As public facilities always tell the truth regarding treatment progress and the probability distribution of D is symmetric, prices in the public sector are always set under the expectation that $A = 0$. In the case of private provision, A is known in advance and affects prices. If $A = -1$ (but $v = 0$), we get

$$p_o^*(t_q) = \frac{c + C(t(\widetilde{\alpha}(t_q))) - S(t_c(\widetilde{\alpha}(t_q)))}{\beta' - \gamma}.$$

If γ is sufficiently large, we get that $p_o^*(t_q) > p_m(t_q)$ for all $t_q \in [0, t^*]$, implying that Proposition 3 does not hold. If $A = 1$, $p_o^*(t_q)$ instead falls compared to $A = 0$. Hence, $p_o^*(0)$ falls and the cutoffs $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ for each level of quality, implying that each level of quality is feasible for a larger range of values of α and lower prices, though the price for each given α is the same.

A.8. Proof of Proposition 4. As the manager has a direct concern for truth-telling and no financial stake in the social service's decision whether to shorten or prolong treatment, he just sets $I = D$. Knowing this, the social service sets $A = I$ and $A = I = D$ is the only equilibrium under public provision. As the naive social service believes that the private provider will set $I = D$, they set $A = I$ also in this case. A sophisticated social service does not believe the signals of private facilities, but since there is a positive probability that a social services is naive, private providers set $A = 1$.

TABLE A1. Summary statistics of variables, data at the placement level

	Public				Private				
Variable	Obs	Min	Max	Mean	Obs	Min	Max	Mean	p
Teenager characteristics									
Previous exp. of residential care*	186	0	1	.296	171	0	1	.444	.004
Previous breakdown*	180	0	1	.211	164	0	1	.402	.000
Psychological disorder*	186	0	1	.177	171	0	1	.269	.037
Addiction*	186	0	1	.269	171	0	1	.415	.004
Evaluated at §12-home*	186	0	1	.059	169	0	1	.107	.104
Non-voluntary placement*	186	0	1	.231	170	0	1	.353	.011
Violent behavior*	186	0	1	.167	171	0	1	.222	.184
Criminal behavior*	186	0	1	.333	171	0	1	.468	.010
Sex* (male = 1)	186	0	1	.484	171	0	1	.556	.176
Immigrant background*	186	0	1	.392	171	0	1	.351	.417
Age* (1991)	186	13	16	14.8	171	13	16	14.9	.360
Problems at home during childhood*	186	0	1	.527	171	0	1	.614	.097
Placed in care due to own behavior*	186	0	1	.538	171	0	1	.684	.005
Previously in youth- or foster care*	183	0	1	.240	169	0	1	.456	.000
Placement number	186	1	4	1.14	171	1	4	1.30	.004

The p-value regards the hypothesis that the proportion or mean of a certain variable is equal in private and public facilities. Variables marked with an asterisk are included in the vector of basic teenager characteristics.

TABLE A1, cont.

Variable	Public				Private				p-value
	Obs	Min	Max	Mean	Obs	Min	Max	Mean	
Facility characteristics									
CAB facility (yes = 1)	181	0	1	.735	–	0	0	–	–
Nonprofit facility (yes = 1)	–	0	0	–	157	0	1	.325	–
Distance (>100 km = 1)*	186	0	1	.081	171	0	1	.468	.000
Places*	185	4	46	9.0	157	2	46	10.2	.204
Personnel density	185	.25	5.00	1.49	154	.14	3.08	.87	.000
Outsider: Municipality (yes = 1)	186	0	1	.016	–	–	–	–	–
Outsider: CAB (yes = 1)	186	0	1	.054	–	–	–	–	–
Treatment: Evaluation* (yes = 1)	183	0	1	.306	155	0	1	.052	.000
Treatment: School* (yes = 1)	183	0	1	.279	155	0	1	.277	.979
Treatment: Work (yes = 1)	183	0	1	.005	155	0	1	.084	.000
Treatment: Family (yes = 1)	183	0	0	.000	155	0	1	.168	.000
Treatment: Milieu therapy (yes = 1)	183	0	1	.344	155	0	1	.174	.000
Treatment: Other (yes = 1)	183	0	1	.148	155	0	1	.232	.046
Treatment: Not stated (yes = 1)	186	0	1	.016	163	0	1	.049	.079
Treatment: "Medlevarskap" (yes = 1)	183	0	1	.044	155	0	1	.187	.000

The p-value regards the hypothesis that the proportion or mean of a certain variable is equal in private and public facilities. Variables marked with an asterisk are included in the vector of basic facility characteristics.

TABLE A1, cont.

Variable	Public				Private				p-value
	Obs	Min	Max	Mean	Obs	Min	Max	Mean	
Municipality characteristics									
Log (poulation 1990)*	175	8.60	13.42	11.42	150	8.72	13.42	11.34	.564
Right-wing majority in municip. legislature* (yes = 1)	173	0	1	.399	145	0	1	.338	.263
Number of teenagers placed in youth care 1991	175	13.086	1	40	150	1	40	10.47	.105
Geography: "Norrland"* (yes = 1)	175	0	1	.097	150	0	1	.120	.508
Geography: "Svealand"* (yes = 1)	175	0	1	.371	150	0	1	.493	.027
Geography: "Götaland"* (yes = 1)	175	0	1	.531	150	0	1	.387	.009
Geography: "Stockholm"* (yes = 1)	175	0	1	.246	150	0	1	.247	.984
Geography: "Göteborg"* (yes = 1)	175	0	1	.149	150	0	1	.053	.005
Geography: "Malmö"* (yes = 1)	175	0	1	.034	150	0	1	.040	.785
Outcome measures									
Cost	172	18,984	92,000	50,645	149	19,410	140,300	45,129	.004
Breakdown of placement (yes = 1)	186	0	1	.258	171	0	1	.392	.007
Treatment length (months)	185	1	41	10.11	170	0	74	20.78	.000
Treatment continued (yes = 1)	179	0	1	.458	153	0	1	.340	.029
Multiple placements 1991	172	0	1	.302	171	0	1	.298	.697
Post-treatment: Social assistance (yes = 1)	136	0	1	.257	142	0	1	.246	.835
Post-treatment: At most primary school (yes = 1)	134	0	1	.649	141	0	1	.660	.857
Post-treatment: Convicted (yes = 1)	136	0	1	.485	142	0	1	.521	.550
Post-treatment: Imprisoned (yes = 1)	136	0	1	.169	142	0	1	.197	.546
Post-treatment: Mental health problems (yes = 1)	136	0	1	.228	142	0	1	.254	.618

The p-value regards the hypothesis that the proportion or mean of a certain variable is equal in private and public facilities.
Variables marked with an asterisk are included in the vector of basic municipality characteristics.

TABLE A2. Robustness checks, cost

	Price (SEK)		
	(1) OLS	(2) OLS	(3) OLS
Constant	51,360***	47,464***	20,086
	(15,969)	(12,108)	(16,960)
Private	−25,742**	−14,927*	−16,465**
	(10,011)	(7,638)	(7,774)
Personnel density	2,860	9,726**	10,540***
	(6,953)	(4,030)	(3,922)
Private*Personnel density	28,052***	21,696***	21,675***
	(7,697)	(5,732)	(5,652)
Nonprofit	−3,201	1,619	855
	(6,908)	(4,672)	(4,379)
CAB	−19,123	−3,549	981
	(17,982)	(4,598)	(4,593)
Nonprofit*Personnel density	4,454		
	(9,056)		
CAB*Personnel density	12,903		
	(12,107)		
Private*Right wing		−10,982**	
		(4,827)	
Private*Log(pop. 1990)			−4,311**
			(1,938)
Teenager characteristics	Yes	Yes	Yes
Facility characteristics	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes
Municipality FE	No	No	No
N	258	258	258
R^2	0.58	0.59	0.59

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE A3. Post-treatment outcomes I

	Treatment continued (APE)					Mental health (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) OLS	(6) Probit	(7) Probit	(8) Probit	(9) OLS
Constant					0.254				0.618
					(0.230)				(0.420)
Private	−0.109	−0.163	−0.081	−0.070	−0.012	0.120	0.115	−0.013	0.007
	(0.095)	(0.101)	(0.115)	(0.143)	(0.208)	(0.115)	(0.124)	(0.168)	(0.265)
Nonprofit	−0.007	0.076	0.141	0.127	0.269*	−0.064	−0.095	−0.082	−0.153
	(0.104)	(0.111)	(0.113)	(0.130)	(0.143)	(0.069)	(0.086)	(0.081)	(0.191)
CAB	0.012	0.045	0.031	0.076	0.127	0.096	0.041	0.007	−0.113
	(0.093)	(0.096)	(0.102)	(0.138)	(0.187)	(0.119)	(0.114)	(0.147)	(0.209)
Private*violence							−0.061	−0.091	0.018
							(0.148)	(0.117)	(0.242)
Private*Pr. break							0.018	−0.031	0.076
							(0.143)	(0.128)	(0.241)
Teenager ch.	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes	No	No	No	Yes
N	316	300	295	264	269	260	227	202	211
R^2					0.55				0.61

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE A4. Post-treatment outcomes II

	Social Assistance (APE)				Education (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) OLS	(5) Probit	(6) Probit	(7) Probit	(8) OLS
Constant				−0.161				1.061*
				(0.510)				(0.543)
Private	−0.123	−0.142	−0.122	−0.078	0.001	−0.113	−0.158	−0.417
	(0.075)	(0.095)	(0.155)	(0.247)	(0.102)	(0.133)	(0.148)	(0.280)
Nonprofit	0.090	0.085	0.041	0.128	−0.058	−0.130	−0.143	0.151
	(0.094)	(0.110)	(0.121)	(0.144)	(0.087)	(0.099)	(0.119)	(0.211)
CAB	−0.119*	−0.154*	−0.175	−0.137	−0.061	−0.190*	−0.187	−0.298
	(0.068)	(0.085)	(0.135)	(0.238)	(0.099)	(0.113)	(0.133)	(0.260)
Private*violence		−0.110	−0.094	−0.256		−0.030	0.068	0.205
		(0.123)	(0.131)	(0.216)		(0.167)	(0.174)	(0.318)
Private*Pr. break		0.054	0.145	0.250		−0.289	−0.238	0.254
		(0.147)	(0.188)	(0.288)		(0.197)	(0.218)	(0.288)
Teenager ch.	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	Yes	No	No	No	Yes
N	260	227	207	211	258	225	205	209
R^2				0.61				0.57

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE A5. Post-treatment outcomes III

	Convictions (APE)				Imprisonment (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) OLS	(5) Probit	(6) Probit	(7) Probit	(8) OLS
Constant				0.204				−0.213
				(0.441)				(0.415)
Private	0.024	0.172	0.203	−0.042	0.057	0.050	0.021	0.088
	(0.088)	(0.123)	(0.166)	(0.224)	(0.074)	(0.064)	(0.072)	(0.177)
Nonprofit	0.008	0.004	0.076	−0.019	0.076	0.078	0.074	0.087
	(0.099)	(0.120)	(0.136)	(0.186)	(0.079)	(0.087)	(0.090)	(0.199)
CAB	−0.005	0.013	0.066	−0.294	0.091	0.047	0.005	−0.052
	(0.088)	(0.100)	(0.148)	(0.202)	(0.085)	(0.055)	(0.061)	(0.168)
Private*Violence		−0.335**	−0.303*	−0.395*		−0.067***	−0.067***	−0.458**
		(0.151)	(0.181)	(0.207)		(0.025)	(0.022)	(0.195)
Private*Pr. break		−0.326*	−0.293	−0.219		0.019	0.057	0.072
		(0.172)	(0.198)	(0.299)		(0.072)	(0.108)	(0.259)
Teenager ch.	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	Yes	No	No	No	Yes
N	260	227	207	211	260	227	202	211
R^2				0.71				0.68

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B1. Selection of teenagers I

	Prev. exp.		Prev. break		Psych. pr.		Addiction		Investig. §12	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private	0.223***	0.162*	0.209***	0.124	0.151**	0.124*	0.244**	0.212**	0.142**	0.097*
	(0.081)	(0.084)	(0.071)	(0.076)	(0.077)	(0.072)	(0.104)	(0.107)	(0.061)	(0.051)
Nonprofit	−0.126*	−0.112	−0.056	−0.101	−0.066	−0.023	0.080	0.015	−0.051*	−0.062***
	(0.073)	(0.081)	(0.086)	(0.080)	(0.072)	(0.092)	(0.084)	(0.088)	(0.026)	(0.023)
CAB	0.073	0.078	0.042	0.039	0.059	0.050	0.163	0.164	0.094	0.089*
	(0.081)	(0.076)	(0.078)	(0.070)	(0.077)	(0.065)	(0.113)	(0.111)	(0.067)	(0.053)
Distance		0.146**		0.281***		−0.058		0.151**		0.066*
		(0.074)		(0.076)		(0.056)		(0.066)		(0.039)
Places		−0.001		0.003		−0.003		0.005		0.002
		(0.003)		(0.003)		(0.003)		(0.003)		(0.002)
Treatment: Evaluations		0.029		0.081		−0.086*		−0.014		−0.060***
		(0.067)		(0.074)		(0.051)		(0.076)		(0.021)
Treatment: School		−0.034		0.099		0.097		−0.024		0.017
		(0.073)		(0.071)		(0.065)		(0.066)		(0.030)
N	338	330	326	318	338	330	338	330	337	329

APE from probit regressions. Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B2. Selection of teenagers II

	Voluntary		Sex		Age		Immigrant		v15r	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private	0.157	0.096	0.077	0.019	−0.177	−0.374*	−0.173**	−0.153	0.273***	0.183*
	(0.106)	(0.105)	(0.082)	(0.100)	(0.204)	(0.190)	(0.081)	(0.099)	(0.075)	(0.095)
Nonprofit	−0.166**	−0.210***	−0.027	−0.010	0.198	0.188	−0.025	−0.013	−0.115	−0.142*
	(0.073)	(0.063)	(0.097)	(0.106)	(0.195)	(0.214)	(0.080)	(0.092)	(0.078)	(0.081)
CAB	−0.018	−0.004	0.010	0.007	−0.250	−0.234	−0.205***	−0.213***	0.045	0.042
	(0.109)	(0.102)	(0.078)	(0.084)	(0.192)	(0.163)	(0.076)	(0.077)	(0.084)	(0.086)
Distance		0.219***		0.047		0.377**		0.006		0.232***
		(0.079)		(0.071)		(0.154)		(0.071)		(0.073)
Places		0.006		−0.001		0.006		−0.002		0.003
		(0.004)		(0.004)		(0.008)		(0.004)		(0.005)
Treatment: Evaluation		0.015		−0.071		−0.196		0.169**		0.009
		(0.083)		(0.077)		(0.160)		(0.069)		(0.074)
Treatment: School		−0.113		0.092		−0.358**		0.121**		−0.052
		(0.074)		(0.066)		(0.151)		(0.058)		(0.068)
Constant					2.979***	3.042***				
					(0.171)	(0.161)				
N	337	329	338	330	338	330	338	330	335	327
R^2					0.01	0.05				

APE from probit regressions except for Age. Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B3. Selection of teenagers III

	v13fr		Crime		Violent beh.		Childhood pr.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private	0.250***	0.144	0.188*	0.091	0.074	0.028	0.151	0.179
	(0.090)	(0.114)	(0.098)	(0.115)	(0.067)	(0.071)	(0.123)	(0.109)
Nonprofit	0.156	0.115	−0.047	−0.008	−0.019	−0.010	−0.204***	−0.204**
	(0.103)	(0.114)	(0.104)	(0.108)	(0.060)	(0.067)	(0.075)	(0.085)
CAB	0.192**	0.200*	0.036	0.029	0.039	0.038	−0.023	−0.010
	(0.092)	(0.111)	(0.098)	(0.106)	(0.065)	(0.061)	(0.128)	(0.110)
Distance		0.225***		0.171**		0.056		0.025
		(0.068)		(0.081)		(0.057)		(0.074)
Places		0.006		−0.005		−0.000		0.002
		(0.005)		(0.004)		(0.003)		(0.004)
Treatment: Evaluation		−0.181**		−0.067		−0.045		0.113
		(0.089)		(0.089)		(0.049)		(0.089)
Treatment: School		0.070		0.064		0.039		−0.230***
		(0.088)		(0.073)		(0.060)		(0.077)
N	338	330	338	330	338	330	338	330

APE from probit regressions except for Age. Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B4. Selection of municipalities

	All private facilities					Only for-profit facilities				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Right-wing majority	0.009		−0.082			−0.009		−0.064		
	(0.101)		(0.102)			(0.094)		(0.107)		
Log(pop. 1990)	0.096		0.053			0.075		0.027		
	(0.063)		(0.066)			(0.062)		(0.072)		
Placements in 1991	−0.010		−0.002			−0.008		−0.002		
	(0.007)		(0.007)			(0.006)		(0.007)		
Geo: "Göteborg"		−0.195***	−0.263**				−0.122**	−0.143		
		(0.058)	(0.114)				(0.062)	(0.117)		
Geo: "Stockholm"		−0.162**	−0.153				−0.112	−0.073		
		(0.082)	(0.121)				(0.089)	(0.125)		
Geo: "Malmo"		0.081	−0.021				0.159**	0.102		
		(0.063)	(0.098)				(0.071)	(0.123)		
Geo: "Götaland"		−0.095	−0.055				−0.147	−0.120		
		(0.112)	(0.116)				(0.115)	(0.120)		
Geo: "Svealand"		0.114	0.131				−0.001	0.014		
		(0.119)	(0.123)				(0.129)	(0.135)		
Contracting in elderly care 1998				0.003					−0.001	
				(0.003)					(0.004)	
Contracting in child care 1998					0.005					0.002
					(0.005)					(0.006)
N	318	325	318	296	296	261	265	261	241	241
Psuedo R^2	.014	.034	.043	.004	.003	.013	.026	.031	.000	.000

APE from Probit regressions. Robust standard errors with clustering at the municipality level in parentheses. "Norrland" omitted variable. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level. in a two-sided test

TABLE B5. Facilities with a personnel density of 0.8 or more

	Price (SEK)				
Variable	(1)	(2)	(3)	(4)	(5)
Constant	22,348**	18,000*	22,539**	54,571***	25,827**
	(9,841)	(10,308)	(9,705)	(11,042)	(12,248)
Private	−4,348	2,896	−6,966	−18,729	−13,632
	(13,017)	(13,103)	(13,102)	(13,017)	(17,085)
Personnel density	19,994***	24,109***	17,754***	12,421**	13,710
	(6,584)	(6,591)	(6,095)	(5,390)	(8,371)
Private*Pers. density	12,710	8,490	16,860**	22,487**	20,306
	(8,589)	(8,629)	(8,076)	(8,829)	(12,306)
Nonprofit	−1,041	776	4,375	7,460	19,197**
	(6,326)	(6,476)	(6,128)	(7,227)	(8,612)
Teenager ch.	No	Yes	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	Yes
Municipality ch.	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
N	235	226	226	202	206
R^2	0.24	0.29	0.41	0.55	0.84

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B6. Controlling for expected duration of treatment

	Price (SEK)				
	(1)	(2)	(3)	(4)	(5)
Constant	32,539***	34,130***	29,046***	43,062***	25,912**
	(9,290)	(9,102)	(7,770)	(12,686)	(12,058)
Private	−7,142	−830	−8,115	−10,980	252
	(8,993)	(9,132)	(8,474)	(8,074)	(12,515)
Personnel density	15,949***	18,123***	14,523***	11,826***	16,484*
	(5,977)	(5,619)	(4,658)	(3,903)	(8,524)
Private*personnel density	18,956***	16,132**	18,241***	19,798***	15,067*
	(6,948)	(6,780)	(5,567)	(5,815)	(8,903)
Nonprofit	−3,284	−2,081	−240	427	7,692
	(3,964)	(3,857)	(3,873)	(4,713)	(7,418)
Expected duration of treatment	−456**	−731***	−226	−285	−434*
	(200)	(263)	(243)	(278)	(247)
Private*previous breakdown		−1,650	123	426	1,820
		(3,614)	(3,489)	(3,846)	(3,797)
Private*violence		5,150	2,992	1,764	1,131
		(4,782)	(4,570)	(4,772)	(3,860)
Teenager ch.	No	Yes	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	Yes
Municipality ch.	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
N	288	288	288	258	263
R^2	0.40	0.44	0.50	0.57	0.83

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B7. Excluding CAB facilities

	Price (SEK)				
	(1)	(2)	(3)	(4)	(5)
Constant	22,757***	14,554*	32,698**	75,328***	44,864**
	(7,573)	(8,723)	(14,280)	(28,244)	(21,470)
Private	−5,942	−899	−18,547	−22,128	−27,784
	(8,083)	(9,132)	(16,088)	(18,199)	(20,190)
Personnel density	20,320***	24,925***	10,727	8,038	−637
	(4,588)	(5,209)	(12,529)	(16,351)	(18,641)
Private*Pers. dens.	13,004**	7,039	21,869	23,91	34,663*
	(5,574)	(6,409)	(13,424)	(17,240)	(18,533)
Nonprofit	−1,956	−378	1,236	1,67	10,823
	(4,168)	(4,251)	(3,986)	(4,459)	(9,635)
Teenager ch.	No	Yes	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	Yes
Municipality ch.	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
Observations	175	166	165	145	148
R-squared	0.50	0.55	0.61	0.67	0.90

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B8. Controlling for insiders and outsiders

	Price (SEK)				
	(1)	(2)	(3)	(4)	(5)
Constant	29,149***	26,752***	26,749***	37,777***	21,347*
	(8,542)	(9,681)	(8,131)	(13,379)	(12,813)
Private	−12,334	−9,816	−10,36	−13,014	−4,736
	(8,996)	(9,956)	(8,455)	(8,334)	(12,313)
Personnel density	15,477**	17,582***	13,982***	11,045***	15,460*
	(5,936)	(6,505)	(4,808)	(4,192)	(8,895)
Private*Pers. dens	17,847***	15,577**	18,537***	20,224***	16,057*
	(6,725)	(7,447)	(5,690)	(5,907)	(9,194)
Nonprofit	−1,956	−1,595	349	1,206	9,299
	(4,162)	(4,251)	(4,015)	(4,843)	(7,578)
Outsider: Municipality	3,214	4,884	8,053	9,445*	10,672
	(4,755)	(5,159)	(5,991)	(5,152)	(7,336)
Outsider: CAB	2,578	1,916	−318	2,120	−924
	(4,236)	(4,880)	(4,161)	(4,248)	(6,506)
Teenager ch.	No	Yes	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	Yes
Municipality ch.	No	No	No	Yes	No
Municipality FE	No	No	No	No	Yes
Observations	302	289	288	258	263
R-squared	0.36	0.39	0.50	0.57	0.83

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B9. Cost and personnel density in public facilities

	CAB		Municipality	
	Mean	Std	Mean	Std
Cost	51,100	15,040	50,590	10,049
Personnel density	1.505	.274	1.517	.831

TABLE B10. Excluding non-profit facilities

	Treatment breakdown (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) OLS
Private	0.281***	0.288***	0.284**	0.212
	(0.086)	(0.110)	(0.138)	(0.194)
CAB	0.107	0.028	0.001	−0.137
	(0.087)	(0.094)	(0.118)	(0.166)
Private*Violence		−0.275***	−0.230***	−0.463**
		(0.057)	(0.059)	(0.203)
Private*Pr. break		−0.121	−0.158*	−0.187
		(0.100)	(0.089)	(0.213)
Constant				0.304
				(0.469)
Teenager ch.	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No
Municipality FE	No	No	No	Yes
N	287	275	240	244
R^2				0.60

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B11. Breakdowns, control for expected duration

	Treatment breakdown (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) OLS
Constant				0.368
				(0.485)
Private	0.301***	0.256**	0.239*	0.278
	(0.097)	(0.122)	(0.145)	(0.206)
Nonprofit	−0.118	−0.091	−0.082	−0.255**
	(0.072)	(0.076)	(0.080)	(0.129)
CAB	0.093	0.047	0.024	−0.044
	(0.086)	(0.091)	(0.117)	(0.158)
Expected duration	−0.005	−0.000	0.001	−0.002
	(0.004)	(0.005)	(0.007)	(0.009)
Private*Violence		−0.280***	−0.261***	−0.508**
		(0.051)	(0.049)	(0.185)
Private*Pr. break		−0.079	−0.103	−0.071
		(0.101)	(0.107)	(0.196)
Teenager ch.	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No
Municipality FE	No	No	No	Yes
N	313	313	281	286
R^2				0.58

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test.

TABLE B12. Prevalence of problems for violent teenagers

	Public		Private	
Variable	Obs	Mean	Obs	Mean
Previous experience of youth care	31	.323	38	.500
Previous breakdown	30	.367	36	.555
Psychological problems	31	.194	38	.368
Addiction	31	.419	38	.632
Investigated at §12-home	31	.194	37	.216
Non-voluntary placement	31	.258	38	.447
Previously in youth care	31	.323	38	.500
Placed due to own behavior	31	.806	38	.818
Criminal behavior	31	.581	38	.737
Problems at home during childhood	31	.516	38	.658

TABLE B13. Breakdowns on the facility's initiative

	Treatment breakdown (APE)			
	(1) Probit	(2) Probit	(3) Probit	(4) OLS
Constant				−0.181
				(0.311)
Private	0.246***	0.173*	0.137	0.198
	(0.081)	(0.092)	(0.113)	(0.183)
Nonprofit	−0.172***	−0.140***	−0.137***	−0.311**
	(0.040)	(0.041)	(0.040)	(0.156)
CAB	0.067	0.002	−0.042	−0.049
	(0.082)	(0.081)	(0.097)	(0.171)
Private*Violence		−0.155***	−0.136***	−0.443**
		(0.047)	(0.045)	(0.203)
Private*Pr. break		0.093	0.007	0.094
		(0.130)	(0.120)	(0.229)
Teenager ch.	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No
Municipality FE	No	No	No	Yes
N	288	274	245	249
R^2				0.62

The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a one-sided test, except for Nonprofit and CAB which refers to a two-sided test.

TABLE B14. Additional control variables

	Duration (months)			
	(1)	(2)	(3)	(4)
Constant	3.872	9.854	17.704	18.598**
	(11.198)	(6.292)	(12.087)	(7.429)
Private	12.581***	11.962*	14.355***	14.000
	(3.716)	(6.945)	(3.553)	(8.630)
Nonprofit	−3.672	−0.344	−4.398	−1.998
	(3.335)	(4.610)	(3.449)	(4.721)
CAB	0.953	−0.424	2.157	1.592
	(2.626)	(5.573)	(2.937)	(7.741)
Personnel density	2.239	0.805		
	(1.452)	(1.692)		
Breakdown	−8.132***	−9.856***	−7.578***	−9.135***
	(1.771)	(2.673)	(1.961)	(2.739)
Cost (in 1000 SEK)			−0.066	−0.101
			(0.058)	(0.078)
Teenager charact.	Yes	Yes	Yes	Yes
Facility charact.	Yes	Yes	Yes	Yes
Municipality charact.	Yes	No	Yes	No
Municipality FE	No	Yes	No	Yes
Full set of treatment options	Yes	Yes	No	No
Observations	277	282	256	261
R-squared	0.42	0.69	0.40	0.70

Vector of teenager characteristics includes interaction effects between private ownership and troublesome teenagers. The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B15. Excess duration of treatmet

	Duration (months)						
	Original data				Imputed values		
Variable	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Constant	4.640**	4.951**	3.240	−18.783	2.299	3.186	8.372
	(2.026)	(2.091)	(5.975)	(19.197)	(2.246)	(2.168)	(10.155)
Private	8.272**	8.983**	9.466*	−0.985	11.975***	13.822***	14.262***
	(3.661)	(3.627)	(4.944)	(6.014)	(2.216)	(2.162)	(3.130)
Nonprofit	0.228	−0.324	1.652	3.781	−2.690	−3.764*	−4.833*
	(3.582)	(3.492)	(4.563)	(5.493)	(2.175)	(2.104)	(2.478)
CAB	−0.529	0.205	2.185	−4.889	1.454	2.124	1.800
	(1.908)	(2.035)	(2.347)	(4.914)	(2.149)	(2.073)	(2.794)
Breakdown		−5.103*	−5.998**	−6.855**		−7.708***	−7.488***
		(2.721)	(2.953)	(2.976)		(1.466)	(1.716)
Planned dur.	0.516***	0.552***	0.420***	0.366***	.546***	.589***	.371***
	(0.118)	(0.115)	(0.105)	(0.103)	(.104)	(.101)	(.108)
Teenager ch.	No	No	Yes	Yes	No	No	Yes
Facility ch.	No	No	Yes	Yes	No	No	Yes
Munic. ch.	No	No	No	Yes	No	No	Yes
Munic. FE	No	No	No	No	No	No	No
$p: \widehat{\beta}_2 - \widehat{\beta}_5$					.00	.00	.00
N	103	103	95	86	341	341	284
R^2	0.36	0.39	0.59	0.65	.21	.27	.42

Vector of teenager characteristics includes interaction effects between private ownership and troublesome teenagers. The standard errors within parentheses have been corrected for clustering at the facility level and heteroskedasticity for (1) to (4), but only not clustered for (5) to (8). One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B16. PT outcomes I, teenagers with one placement

	Social assistance				Education			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant				−0.165				1.102**
				(0.513)				(0.539)
Private	−0.111	−0.227**	−0.162	−0.107	−0.005	−0.204	−0.169	−0.447
	(0.080)	(0.104)	(0.162)	(0.287)	(0.102)	(0.142)	(0.156)	(0.279)
Nonprofit	0.075	0.063	0.066	0.134	−0.005	−0.071	−0.060	0.320
	(0.101)	(0.115)	(0.130)	(0.138)	(0.099)	(0.123)	(0.130)	(0.199)
CAB	−0.125*	−0.162*	−0.156	−0.040	−0.049	−0.185	−0.144	−0.249
	(0.071)	(0.095)	(0.148)	(0.275)	(0.095)	(0.118)	(0.138)	(0.250)
Private*Violence	0.045	−0.025	−0.192		0.110	0.126	0.208	
		(0.192)	(0.164)	(0.227)		(0.158)	(0.163)	(0.334)
Private*Pr. break	0.100	0.111	−0.023		−0.337	−0.310	0.372	
		(0.199)	(0.216)	(0.383)		(0.238)	(0.244)	(0.306)
Teenager ch.	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	Yes	No	No	No	Yes
N	224	195	191	195	223	194	190	194
R^2				0.66				0.62

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B17. PT outcomes II, teenagers with one placement

	Convictions			Imprisonment				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant				0.263				−0.132
				(0.416)				(0.427)
Private	0.045	0.115	0.229	−0.038	0.065	0.066	0.020	0.033
	(0.098)	(0.141)	(0.168)	(0.274)	(0.084)	(0.086)	(0.068)	(0.202)
Nonprofit	0.049	0.066	0.074	−0.041	0.084	0.028	0.038	0.109
	(0.116)	(0.134)	(0.145)	(0.229)	(0.093)	(0.075)	(0.067)	(0.202)
CAB	0.024	0.035	0.132	−0.246	0.090	0.068	0.005	−0.116
	(0.095)	(0.111)	(0.161)	(0.248)	(0.085)	(0.067)	(0.056)	(0.208)
privateviolence	−0.311*	−0.226	−0.351*			−0.075***	−0.053**	−0.484**
		(0.167)	(0.202)	(0.201)		(0.028)	(0.023)	(0.238)
Private*Pr. break	−0.381**	−0.410**	−0.407			−0.034	0.031	−0.028
		(0.187)	(0.174)	(0.307)		(0.051)	(0.093)	(0.315)
Teenager ch.	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	Yes	No	No	No	Yes
N	224	195	191	195	224	195	187	195
R^2				0.72				0.69

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B18. PT outcomes. III, one placement

	Mental health			
	(1)	(2)	(3)	(4)
Constant				0.593
				(0.399)
Private	0.047	0.115	−0.105	−0.170
	(0.114)	(0.124)	(0.163)	(0.225)
Nonprofit	0.012	−0.051	−0.094	−0.101
	(0.082)	(0.101)	(0.084)	(0.194)
CAB	0.093	0.060	−0.086	−0.279
	(0.113)	(0.108)	(0.143)	(0.196)
Private*Violence		−0.118	−0.053	0.054
		(0.117)	(0.141)	(0.265)
Private*Pr. break		−0.096	−0.092	−0.017
		(0.123)	(0.118)	(0.333)
Teenager ch.	No	Yes	Yes	Yes
Facility ch.	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No
Municipality FE	No	No	No	Yes
N	224	195	187	195
R				0.66

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

TABLE B19. Post-treatment outcomes, troublesome teenagers

	Convictions				Imprisonment			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Constant				1.483***				0.222
				(0.337)				(0.335)
Private	−0.340**	−0.398***	−0.152	−0.470	−0.087	−0.035	−0.348*	−0.220
	(0.170)	(0.145)	(0.229)	(0.322)	(0.172)	(0.076)	(0.209)	(0.322)
Nonprofit	−0.029	−0.124	0.039	−0.088	0.135	0.104	0.346*	0.201
	(0.128)	(0.164)	(0.128)	(0.396)	(0.154)	(0.113)	(0.196)	(0.437)
CAB	−0.220	−0.332	−0.004	−0.181	0.091	0.059	−0.176	−0.048
	(0.213)	(0.213)	(0.233)	(0.294)	(0.183)	(0.097)	(0.195)	(0.332)
Teenager ch.	No	Yes	No	No	No	Yes	No	No
Facility ch.	No	No	Yes	Yes	No	No	Yes	Yes
Municipality ch.	No	No	Yes	No	No	No	Yes	No
Municipality FE	No	No	No	Yes	No	No	No	Yes
N	96	84	79	83	96	86	79	83
R^2				0.68				0.65

Robust standard errors clustered on the facility level in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level in a two-sided test.

References

Andreassen, T. (2003), *Institutionsbehandling av ungdomar – Vad säger forskningen?,* Centrum för utvärdering av socialt arbete och Statens institutionsstyrelse.

Bayer, P. and Pozen, D. E. (2005), "The Effectiveness of Juvenile Correctional Facilities: Public Versus Private Management", *Journal of Law and Economics*, October.

Gibbs, I. and Sinclair, I. (1998), "Private and Local Authority Children's Homes: A Comparison", *Journal of Adolescence*, 21, 517–527.

Haglund, A., Andersson, B. and Cinandler, R. (1993), *Ekonomistyrning i förändring – kommunal förnyelse och effektivisering*, Kommundata AB och Kommentus Förlag AB, Älvsjö.

Hart, O., Shleifer, A. and Vishny, R. W. (1997), "The Proper Scope of Government: Theory and an Application to Prisons", *Quartely Journal of Economics*, November 1997, Vol. 112, 1127–1161.

Holmström, B. and Milgrom, P. (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics and Organization*, Vol 7, 24–51.

Lindqvist, E. (2007), "Will Privatization Reduce Costs?", SSE/EFI Working Paper Series in Economics and Finance No. 660.

Lipsey, M. W. (1999), "Can Intervention Rehabilitate Serious Delinquents?", *Annals of the American Academy of Political and Social Sciences*, Vol. 64, 142–166.

National Board of Health and Welfare (1990), *Vård utom hemmet* (SoS-rapport 1990:4), NBHW, Stockholm.

National Board of Health and Welfare (1991), *Privat och offentligt i vården* (SoS-rapport 1991:27), NBHW, Stockholm.

National Board of Health and Welfare (1991), *Förteckning över hem för vård eller boende samt familjevårdsenheter 1991*, NBHW, Stockholm.

National Board of Health and Welfare (1993), *Förteckning över hem för vård eller boende samt familjevårdsenheter 1993*, NBHW, Stockholm.

National Board of Health and Welfare (1994), *Barn- och ungdomsinstitutioner – Hur påverkas vården av kommunernas ekonomiska situation?* (SoS-rapport 1994:10), NBHW, Stockholm.

National Board of Health and Welfare (2001), *Barn och unga - insatser 2000* (SoS-rapport 2001:8), NBHW, Stockholm.

Parliamentary Auditors (2002), *Med tvång och god vilja – vad gör Statens institutionsstyrelse?*, Parliamentary Auditors (2002/03:1), Stockholm.

Sallnäs, M. (2000), *Barnavårdens institutioner – framväxt, ideologi och struktur*, Socialhögskolan, Stockholms universitet, Stockholm.

Sallnäs, M. (2005), "Vårdmarknad med svårigheter – om privata aktörer inom institutionsvården för barn och ungdomar", *Socialvetenskaplig tidskrift*, No. 2–3, 226–245.

Schmidt, K. M. (1996), "The Costs and Benefits of Privatization: An Incomplete Contracts Approach", *Journal of Law, Economics and Organization*, Vol. 12, No. 1, 1–24.

Shleifer, A. (1998), "State versus Private Ownership", *Journal of Economic Perspectives*, Fall 1998.

Swedish National Audit Office (2002), *Tillsyn av behandlingshem för barn och ungdomar* (Riksrevisionsverket rapport 2002:6), Swedish National Audit Office, Stockholm.

Vinnerljung, B., Sallnäs, M. and Kyhle-Westermark, P. (2001), *Sammanbrott vid tonårsplaceringar – om ungdomar i fosterhem och på institution*, CUS – Socialstyrelsen, Stockholm.

PAPER 3

Identity and Redistribution

Erik Lindqvist and Robert Östling

ABSTRACT. This paper models the interaction between individuals' identity choices and redistribution. Both redistributive polices and identity choices are endogenous, and there might be multiple equilibria. The model is applied to ethnicity and social class. In an equilibrium with high taxes, the poor identify as poor and favor high taxes. In an equilibrium with low taxes, at least some of the poor identify with their ethnic group and favor low taxes. The model has two main predictions. First, redistribution is highest when society is ethnically homogenous, but the effect of ethnic diversity on redistribution is not necessarily monotonic. Second, when income inequality is low, an increase in income inequality might induce the poor to identify with their ethnic group and therefore favor lower taxes.

1. Introduction

> The Marxian solidarity between the toilers of all the earth will, indeed, have a long way to go as far as concerns solidarity of the poor white Americans with the toiling Negro. (Myrdal 1944, p. 69)

Both the canonical economic theory of redistribution (e.g. Meltzer and Richard 1981) and Marxian theory assume that people's political preferences are determined by their economic position in society. This view is controversial. Conflicts along other dimensions, such as ethnicity, race, religion or gender, may be more important than social class. In particular, it has often been argued that class conflict is rare in societies that are ethnically divided. For example, the racial diversity among the American working class is a recurring theme in the literature on the failure to establish a strong worker's movement in the United States.[1]

[0] We would like to thank Alberto Alesina, Johan Almenberg, Kaushik Basu, Milo Bianchi, Tore Ellingsen, Ola Granström, Magnus Johanneson, Glenn C. Loury, Erik Mohlin, Torsten Persson, Daniel N. Posner, Per Sonnerby, Björn Tyrefors, Björn Wallace and seminar participants at the EEA Congress 2006, Massachusetts Institute of Technology, Stockholm School of Economics and the Zeuthen Workshop on Behavioral Economics 2005 for valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[1] See, for example, Myrdal (1944), Glazer and Moynihan (1970) and Lipset and Marks (2000).

The view that there are multiple dimensions of political conflict invokes the questions under what circumstances voters come to identify with a certain group and how their identities affect redistributive policies. Social psychology research indicates that people tend to identify with groups that have high status, suggesting that redistribution in turn might affect identity choices.[2] In this paper, we develop a formal framework where both redistribution and identity choices are endogenously determined. We use ethnicity and social class (defined as intervals of the income distribution) as our leading example throughout the paper, but the model is applicable to any situation where there are two potential dimensions of social cleavage.[3]

We view identity as altruism toward a subset of the population. Each agent belongs to an ethnic group and a social class, but chooses to identify with only one of them. Two factors determine the identity choice. First, agents want to identify with groups with high status, which is given by the group's average after-tax income. Second, identification with a group involves a cognitive cost which is determined by the proportion of different types in the group. For example, a person feels more distant to her social class if it consists of many people from other ethnic groups. Identity choices affect preferences for redistribution, and redistribution in turn affect identity choices through its effect on the status of different identities. In equilibrium, we require that voting and identity choices are consistent in the sense that nobody wants to switch identity or vote differently. This equilibrium concept is called *Social Identity Equilibrium* and is due to Shayo (2005). We show that there always exists at least one equilibrium in the general case with any number of ethnic groups and social classes. In addition, we show that in this general setting, but with restrictive assumptions on the income distribution, increasing the size of a small ethnic group, or adding a small ethnic group, leads to lower or unchanged levels of redistribution. This is in line with the empirical finding that ethnic diversity is associated with lower levels of redistribution (see Alesina and La Ferrara 2005 for a survey of this literature).

In order to make the analysis more tractable, we further specify the model so that there are only two income levels (rich and poor) and two ethnic groups (black and white). We assume that the poor are in majority, blacks are in minority and the average income of blacks is the same or lower as the average income of whites. In this setting, the level of redistribution is determined by the identity choices of poor blacks and poor whites.

---

[2] See the study by Roccas (2003) and references cited therein for empirical evidence that people tend to identify with groups that have high status.

[3] See Posner (2005) for a variety of different examples of two-dimensional social cleavages, and Alesina and La Ferrara (2005*a*) for further motivation why ethnic identitification is likely to be endogenous to economic policy.

Poor blacks and poor whites choose whether to identify with the poor or their respective ethnic group. Identity choices are determined by a trade-off between the relatively higher status of the ethnic identities, and the potentially lower cognitive distance to the poor identity. Identity choices in turn affect voters' preferences for redistribution. If poor blacks and poor whites identify as poor, they are altruistic toward the poor and vote for a high tax rate. If they instead identify with their respective ethnic groups, they favor low taxes since their altruism is now confined to the relatively richer black and white groups. The implemented tax rate in turn determines the status of the poor, black and white identities through the effect on the average after-tax income of these groups.

Poor whites are most prone to identify as poor and favor high taxes when there are no blacks in society at all. As the number of blacks increases, the cognitive distance to the poor identity grows and poor whites might switch to a white identity and favor lower taxes. This mechanism can explain why social class seems to be more important, and redistribution higher, in ethnically more homogeneous societies (e.g. Scandinavia compared to the US). However, for poor blacks, a higher proportion of blacks also reduces the cognitive distance to the poor identity. Consequently, an increase in the number of blacks might induce poor blacks to identify with the poor and favor more redistribution. The effect of ethnic diversity on redistribution is therefore not necessarily monotonic in our model.

The standard model of redistribution (e.g. Meltzer and Richarcd 1981) predicts that an increase in pre-tax income inequality, in the sense of a larger distance between the median and average income, leads to more redistribution. In contrast, but in accordance with empirical evidence (e.g. Perotti 1996 and Lind 2005), higher pre-tax income inequality need not imply more redistribution in our model. The reason is that higher income inequality may increase the status difference between ethnic and poor identities, which makes the poor more likely to identify with their ethnic groups and favor low taxes. The model also allows us to study the effect of a change in income differences across ethnic groups. An increase in interethnic income inequality changes the cognitive distance to both ethnic and poor identities as well as preferred tax rates for given identity choices. The main prediction is that the level of redistribution falls as interethnic income inequality increases if the black minority group is small, but might increase if the black group is large.

We also study more complicated interactions between ethnic diversity, income inequality and redistribution. Consider for example an inflow of poor black immigrants. Such an inflow has counteracting effects on the identity choices of poor blacks. On the one hand, both the status of the black identity and the cognitive distance to the

poor identity decrease, which makes it more likely that poor blacks identify as poor and favor high taxes. On the other hand, a higher proportion of poor among blacks also implies that the cognitive distance to the black identity decreases. If there are few black people, the latter effect is stronger and an increase in the proportion of poor blacks might induce poor blacks to identify with the black group and favor less redistribution. Somewhat paradoxically, an increase in the proportion of poor blacks may therefore reduce the support for redistribution also among poor blacks.

A feature of the model is that there is a complementarity between tax rates and identity choices. A higher tax rate increases the status of the poor identity relative to ethnic identities and makes it more likely that poor whites and poor blacks identify with the poor and prefer higher taxes. The complementarity between tax rates and identity choices implies that there might be multiple equilibria. For example, we may have one high tax equilibrium where the poor identify as poor and one low tax equilibrium where they identify with their respective ethnic group.

The richness of the predictions and the presence of multiple equilibria suggests that the model may be difficult to test empirically. However, all but one (Proposition 6) of the formally stated results provide predictions given initial identity choices, incomes and population proportions of different groups. Income and population proportions are easily available data, and there are several ways to empirically measure people's identities, for example using survey responses or the probability of homogamy (see Bisin et al. 2006 for a recent example). Given such data, our model provides empirically testable predictions for both the level of redistribution and individuals' identity choices.

Our model extends the redistribution model in Shayo (2005), where individuals have the choice between identifying with their social class (rich or poor) and a common nationalist identity. There are two main conceptual differences. First, extending the model with an ethnic dimension, i.e., several ethnic identities rather than a single nationalist identity, implies that more than one group can influence the tax rate. This creates an interdependence between the identity choices of different groups. Second, we define individuals' cognitive distances to different identities as a function of the proportions of different types in the population. This allows us to explicitly study the effects of changes in the demographic composition of society, such as an increase in ethnic diversity. These differences affect the results. For example, Shayo (2005) argues that an increase in ethnic diversity concentrated among the poor reduces redistribution, which is not always the case in our model.

Our approach differs from previous economic theories of ethnic diversity and redistributive policies.[4] First, there are models that expand the policy space with a non-economic issue or targeted transfers. Roemer (1998) studies how an additional non-economic political issue such as religion or race leads to a bundling effect of political policies.[5] A citizen that favors a high degree of redistribution may vote for a political party that advocates a low degree of redistribution if he favors the political party's position on racial issues. Fernández and Levy (2007) instead consider endogenous political parties where the policy space consists of general redistribution and targeted public goods. For intermediate levels of preference diversity they find that the rich might form a winning coalition with special interest groups among the poor to reduce general redistribution. In contrast to these models, we consider a one-dimensional policy space. A second type of explanation, put forward by, e.g., Alesina et al. (1999) and Alesina et al. (2001), concerns a direct effect of ethnic fragmentation on voter preferences for redistribution. In their view, a voter's altruistic motive for redistribution is confined to people that belong to her own ethnic group. Common to both types of explanations is that voters' political preferences on ethnic issues are exogenous, whereas both preferences and redistribution are determined endogenously in our model.

In the next section of the paper, we develop the model with arbitrarily many ethnic groups and social classes. In Section 3 we specify the black and white model that is used throughout the rest of the paper. The following three sections discuss the implications of the black and white model. Section 7 discusses how "American Exceptionalism" – the difference in redistribution between the United States and Western Europe – can be explained by our model. Section 8 concludes the paper.

2. General Model

In this section we extend the model of redistribution in Shayo (2005) to arbitrarily many ethnic and class identities.[6] The model could equally well be applied to identities along two other dimensions, for example language and religion.

[4] More generally, ethnic heterogeneity might of course also influence economic outcomes through other channels than the political system. For example, ethnicity might influence the ease by which people cooperate, act as focal points in coordination games or affect the possibility to enforce social forms through social networks. See Habyarimana et al. (2007) for references and an overview of this literature.

[5] Austen-Smith and Wallerstein (2003) develops a related model of legislative bargaining.

[6] We refer to Shayo (2005) for a detailed discussion of how the model relates to the social psychological research on social identity.

Consider a set of N agents, a finite set C of social classes and a finite set E of ethnic groups. We view a social class as a particular interval of the income distribution, i.e., all agents within a certain income interval belongs to the social class corresponding to that interval.[7] Each agent also belongs to an ethnic group.[8] All social classes are represented in every ethnic group, and we refer to a particular combination of class and ethnicity as a type. Agents must choose to identify with either their ethnic group or their social class.[9] Given this identity choice, agents also choose which tax rate to vote for. Simple majority voting selects the winning tax rate and in equilibrium we require that the resulting tax rate is consistent with identity and voting choices.

Each agent in the economy is endowed with pre-tax income $y_i > 0$ and the average income in the population is denoted by y. There is a single proportional tax rate t. Tax revenues are redistributed lump-sum and there is a quadratic deadweight loss of taxation equal to $(t^2/2)\,y$.[10] This implies that that the income after taxes and transfers of agent i is $\bar{y}_i = (1-t)\,y_i + (t - t^2/2)\,y$. Similarly, let y_j denote the average pre-tax income of the agents belonging to ethnic group $j \in E$ or social class $j \in C$ so that their average after-tax income (including transfers) is given by $\bar{y}_j = (1-t)\,y_j + (t - t^2/2)\,y$.

Since each agent belongs to one social class and one ethnic group, the average income of these two categories will generally differ. For an agent with low income, the average income in her ethnic group will typically be higher than in her social class, whereas for rich people the social class will typically have a higher average income than the ethnic group. We refer to the category with the higher pre-tax income as the agent's high status identity and the other as the low status identity. The average pre-tax income in the high status versus low status identity is denoted by y_H and y_L

[7] It is not necessary to define social classes in terms of income intervals – one can think of more complicated mappings that takes educational and cultural aspects into account. Proposition 1 holds also with such alternative interpretations.

[8] We assume that there is an uncontroversial way to determine which ethnic group each agent in the economy belongs to. In practice, this is of course easier said than done. For an axiomatic approach to determination of group membership, see Kasher and Rubinstein (1997).

[9] This assumption raises three related issues. First, why can't an individual identify with her type, i.e., her particular combination of class and ethnicity? We don't allow this since it would be very similar to allowing each agent to identify with herself, which corresponds to the standard model of redistribution. Second, why can't an individual identify with both her ethnic group and her social class? In this setting, an agent "can't have both" since she has to vote for her preferred level of redistribution which forces her to decide on how much to favor either of her two groups. However, it is straightforward to allow for intermediate identification, e.g. 30 percent class identification and 70 percent ethnic identification (see footnote 11). Third, we don't allow agents to identify with a group they don't belong to. Though we could allow agents to identify with any group in society, this aspect is not relevant in contexts where it is very costly to shift ethnic identity (for example, from black to white in the US).

[10] The results in this section of the paper holds as long as the deadweight loss is strictly convex in t so that unique solutions to agents' voting problems exist and preferences are single-peaked.

where $y_H \geq y_L$. The identity choice consists of choosing l_i to be either zero or one, where $l_i = 1$ means that the agent identifies with the low status group and $l_i = 0$ that she identifies with her high status group.[11]

An agent's utility consists of two parts: material utility arising from after-tax income including transfers and the immaterial utility arising from identification with a group. The utility function is assumed to be additively separable and take the following form:

$$U_i = \bar{y}_i(t) + \gamma\left(l_i \bar{y}_L(t) + (1 - l_i)\,\bar{y}_H(t)\right) - \delta\left(l_i d_L + (1 - l_i)\,d_H\right), \tag{2.1}$$

where t is the prevailing tax rate and $\bar{y}_H(t)$ and $\bar{y}_L(t)$ are the after-tax incomes of the two categories the agent belongs to. The first term in the utility function represents direct material benefit of after-tax income, the second term the status of the group the individual identifies with, and the last term the cognitive distance to the same group. The parameters γ and δ are positive so that utility is increasing in status and decreasing in cognitive distance. Group status is linearly increasing in the group's after-tax average income and so the higher income group has higher status than the lower income group.

The cognitive distance to a group is higher if there are many people of a different type than oneself in the group. To make this precise, let p_{jk} denote the proportion in the population that belong to social class $j \in C$ and ethnic group $k \in E$. The distance of an agent that belongs to social class j and ethnic group k to his social class j is given by

$$d_{jkj} = d\left(\beta \sum_{h \in E \setminus \{k\}} p_{jh} / \sum_{h \in E} p_{jh}\right),$$

where $d(\cdot)$ is some positive and increasing real-valued function and $\beta > 0$. In other words, the distance to the class identity is increasing in the proportion of people belonging to the same social class that is from a different ethnic group. The parameter β is a measure of ethnic tensions – if β is high, the distance to the class identity is large since the members of a social class come from different ethnic groups. Similarly, an agent that belongs to social class j and ethnic group k has the following distance to his ethnic group k:

$$d_{jkk} = d\left(\alpha \sum_{h \in C \setminus \{j\}} p_{hk} / \sum_{h \in C} p_{hk}\right),$$

[11] We assume that an agent cannot partially identify with a group. However, Proposition 1 is unaffected if the agent is allowed to pick $l_i \in [0, 1]$ as long as the specification implies a unique solution $l_i^*(t)$ that is non-decreasing in t.

where $\alpha > 0$ is a measure of class tensions. Note that cognitive distances do not depend on tax rates – the tax rate only affects the material utility and the relative status of groups. The above specifications imply that the distance to an identity is unaffected by the identity choice of other agents – the distance to a certain identity only depends on characteristics of the population.

The tax rate t is determined by simple majority voting or some other political process that selects the median tax under the assumption of single-peaked preferences. The political process will hence be a mapping Γ from the vector of all tax votes $t^* \in \times_{i \in N}[0,1]$ to a median tax rate $t \in [0,1]$.

Following the definition of social identity equilibrium in Shayo (2005) we require that three conditions must hold in equilibrium:

(1) All individuals vote for their preferred tax rate given the identity choice l_i:

$$t_i^*(l_i) = \arg\max_{t \in [0,1]} \{\bar{y}_i(t) + \gamma(l_i \bar{y}_L(t) + (1 - l_i)\bar{y}_H(t)) - \delta(l_i d_L + (1 - l_i) d_H)\}.$$

(2) All agents choose identity optimally given the prevailing tax rate t:

$$l_i^*(t) = \arg\max_{l_i \in \{0,1\}} \{\bar{y}_i(t) + \gamma(l_i \bar{y}_L(t) + (1 - l_i)\bar{y}_H(t)) - \delta(l_i d_L + (1 - l_i) d_H)\}.$$

(3) The median tax rate is consistent with identity choices and voting behavior of all individuals:

$$\Gamma(t^*(l^*(t))) = t.$$

Note that agents do not choose identities and tax rates simultaneously. The main reason for this assumption is that these are two conceptually different decisions that are likely to be made under different circumstances and at different points in time.[12] A second reason is that preferences for taxes are single-peaked only for given identity choices. In order to be able to use the median voter theorem we cannot admit agents to switch identity at the same time as they choose tax rates.

First consider the agents' voting choices. The utility function (2.1) is strictly concave in t and we can therefore use the first-order condition to derive the solution to the agent's tax voting decision:

[12] Shayo (2005, p. 9) motivates this assumption as follows: "We emphasize that these are equilibrium requirements. We are not asserting that there exists some controlled, deliberative process in which individuals 'choose' their social identities optimally. Rather, we are using the tools of optimization to describe a steady state that takes into account the observed process whereby (a) given cognitive distance, individuals tend to identify with that group that possesses the higher status, and (b) given status, identify with the group more similar to themselves."

$$t_i^*(l_i) = \max\left\{1 - \frac{y_i + \gamma\left(l_i y_L + (1 - l_i)\, y_H\right)}{(1+\gamma)\, y}, 0\right\}. \tag{2.2}$$

Note that this tax rate is non-decreasing in l_i, i.e., the more the agent identifies with the low status identity, the higher is her preferred tax. The reason is that people are altruistic toward the group they identify with. Since the low status group is poorer, an agent favors more redistribution if she identifies with that group.

Now consider optimal identity choices. For a given tax rate t, an agent chooses the high status identity, i.e., $l_i = 0$, if[13]

$$\gamma\,(1-t)\,(y_H - y_L) > \delta\,(d_H - d_L)\,. \tag{2.3}$$

It is clear from this condition that $l_i^*(t)$ is non-decreasing in t. In other words, for given cognitive distances, a higher tax rate implies that the low status identity becomes relatively more attractive since redistribution benefits the low status group more. The higher the prevailing tax rate, the more likely it is that people identify with their low status identity, which in turn would imply that they vote for higher tax rates. Since the median tax rate is non-decreasing in the vector of tax rate choices, there is a complementarity between identity choices and the tax rate. This complementarity means that there are potentially many equilibria, but it also allows us to establish that at least one equilibrium exists.

PROPOSITION 1. *There exists at least one social identity equilibrium.*

PROOF. We know that $l_i^*(t)$ is non-decreasing in t and that $t_i^*(l_i)$ is non-decreasing in l_i. Consequently, $t_i^*(l_i^*(t))$ is non-decreasing in t, which implies that the median tax is non-decreasing in t. Equilibrium tax rates are given by the fixed points of $\Gamma(\times_{i\in N} t_i^*(l_i^*(t)))$. Note that this is a non-decreasing mapping $\Gamma(t) : [0,1] \to [0,1]$. This mapping will typically not be continuous, but since it is a non-decreasing mapping from the unit interval to the unit interval, Tarski's fixed point theorem implies that there is at least one fixed point of $\Gamma(t)$ (see Theorem M.I.3 in Mas-Colell et al. 1995). □

It is difficult to derive any general comparative statics without further specifying the model. For example, although we know that an increase in the parameters for class (α) and ethnic (β) tensions would unambiguously affect the identity choice toward identification with the ethnic group or social class, respectively, it is uncertain how this would affect the tax rate since it differs between individuals whether the ethnic

[13] In the unlikely event that an agent is indifferent between the two identities, we will assume that the agent chooses the low status identity.

group or the social class is the high or low status group. Similarly, we can add an ethnic group to the economy so that only the cognitive distance to the class identity increases for the already existing groups. However, the ethnic identity might be either a high or low status identity, which implies that we cannot tell in which direction the tax rate will change. If, in addition, the income of the new group differs from the average income in the population, the voting choices (for given identities) would be affected, which complicates the analysis further. In order to derive results for the effects of an increase in ethnic diversity that are not confounded by income differences between ethnic groups, we first study the simplest possible distribution of income.

Suppose there are only two income levels, $y_R > y_P > 0$, and consequently two social classes, rich (R) and poor (P). The poor are in majority and all ethnic groups have same proportion of poor. From the expression for the most preferred tax rate (2.2) we see that these assumptions imply that the rich prefer zero taxes irrespective of how they identify themselves, and that the poor always prefer positive taxes. The median voter(s) must therefore be poor. From the condition for high status identification (2.3) it is clear that the identity choices of the poor only differ in the distances to the poor identity. The larger the ethnic group a poor individual belongs to, the smaller is the distance to the poor identity and the more likely she is to identify with the poor. There are only two possible equilibrium tax rates in this setting. In the high tax equilibrium, the poor in relatively large ethnic groups identify as poor and they are sufficiently many to create a majority for their preferred tax rate, which is high since they are altruistic toward the poor. In the low tax equilibrium, the poor in relatively small ethnic groups identify with their ethnic groups and are sufficiently many to be pivotal. This tax rate is lower since they are altruistic toward their ethnic groups that contain both rich and poor.[14] Increasing the size of an existing ethnic group, or adding an ethnic group with the same proportion of poor as the already existing ones, implies that all other ethnic groups shrink in relative size, and consequently, that the distances to the poor identity increase. As long as the enlarged or added ethnic group is sufficiently small, the set of parameter values that can support the high tax equilibrium shrinks, which is stated in Proposition 2. Proposition 2 and the propositions that follow below provide conditions under which a change in parameters may render the initial equilibrium unfeasible. Hence, we do not consider the possibility that the economy may shift from one equilibrium to another in the presence of multiple equilibria.

PROPOSITION 2. *If the assumptions in the previous paragraph hold, then increasing the size of an existing ethnic group, or adding a new ethnic group, that has the same*

[14] There are of course also equilibria where all poor identify either as poor or with their ethnic groups.

proportion of poor and is sufficiently small leads to a lower or unchanged equilibrium tax rate. A sufficient condition for the size of the enlarged or added ethnic group is that it is smaller than the largest ethnic group that contain poor that support the initial equilibrium tax rate.

PROOF. It is clear from (2.2) that the poor always prefer positive tax rates, the rich prefer zero taxes, and that preferred tax rates for given identity choices is unchanged. Since the poor are in majority, the median voter(s) is poor. The condition for ethnic identification (2.3) for the poor only differs in the cognitive distances to the poor identity. Letting p_{E_j} denote the proportion of the population belonging to ethnic group E_j, we can re-write the distance to the poor identity for a poor person in this ethnic group as

$$d_{PE_jP} = d\left(\beta\left(1 - p_{E_j}\right)\right).$$

In other words, the higher p_{E_j} is, the lower is the distance to the poor identity and the more likely is identification with the poor. This implies that there is a $\widehat{p}_E$ such that the poor in ethnic groups larger than $\widehat{p}_E$ identify with the poor, and the poor in smaller ethnic groups identify with their ethnic group. It may also be the case that all ethnic groups are above or below this threshold. Increasing the size of one ethnic group or adding a new ethnic group implies that the other ethnic groups shrink in size, which might induce them to shift to an ethnic identity.

Suppose first that the high tax rate prevails in which sufficiently many of the poor identify as poor. Increasing the size of an ethnic group or adding a new ethnic group could then lead some of the poor to switch to ethnic identities, which may result in the low equilibrium tax rate. (The new low equilibrium tax rate may in turn induce the poor in other ethnic groups to switch to ethnic identities.)

Now suppose instead the sufficiently many poor initially identified with their ethnic groups so that the low tax rate preferred by poor identifying with their ethnic group prevails. Enlarging or adding one ethnic group may then induce some of the poor in the other groups to switch to ethnic identities. However, we want to rule out that the enlarged or new group isn't so large that the poor in that group identify with the poor. This cannot happen if the group is smaller than $\widehat{p}_E$. Since this threshold isn't observable, another sufficient condition is that the new or enlarged group is smaller than the largest group that initially identified with their ethnic group. □

In line with Proposition 2, several papers have shown empirically that there is a negative relation between ethnic heterogeneity and redistribution both across countries and between communities within countries. For example, Alesina et al. (2001) found social spending to be lower in countries with a high degree of racial fractionalization,

Alesina et al. (1999) found a lower degree of public goods provision in ethnically fragmented metropolitan areas in the US, and Soss et al. (2001) found that when US states were given greater autonomy to set their own welfare policies, states with higher proportion of blacks implemented more punitive welfare regulations. Luttmer (2001) shows that support for welfare spending in the US is higher among people living in areas where the proportion of welfare recipients from their own racial group is high. Similarly, Orr (1976) found a negative correlation between aid to families with dependent children and the proportion of non-white welfare recipients across US states. However, Dincer and Lambert (2006) found a U-shaped relationship between ethnic fractionalization and redistribution across US states. Our model can accommodate this empirical finding, which is discussed in Section 4.

A seemingly paradoxical finding that our model can explain is why class voting, i.e., the extent to which voting behavior coincide with social class, seems to be particularly important in Scandinavian countries – which have the lowest income inequality in the world. Our answer is that the Scandinavian countries are relatively ethnically homogeneous, suggesting that the poor identify with their class.[15] In line with this explanation, Nieuwbeerta and Ultee (1999) found a negative correlation between religious and ethnic diversity and the level of class voting.

The model also suggests that members of small ethnic groups tend to identify with their ethnicity, which resonates well with the picture of New York in the 1960s described by Glazer and Moynihan (1970).[16] A similar idea has also been used by the authoritarian former leader of Singapore, Lee Kuan Yew, to legitimize Singapore's one-party system:

> In multiracial societies, you don't vote in accordance with your economic interests and social interests, you vote in accordance with race and religion. Supposing I'd run their system [democracy] here: Malays would vote for Muslims, Indians would vote for Indians, the Chinese would vote for Chinese. (Spiegel 2005, p. 23)

Although Lee Kuan Yew may be right that Malays and Indians in Singapore would vote for their own ethnic groups if they were allowed to vote, it is less clear that the Chinese would do so since they constitute roughly three quarters of the population.[17]

---

[15] According to Nieuwbeerta and Ultee (1999), class voting was particularly important in the Scandinavian countries at least until the 1980s. Since then class voting has declined, but on the other hand the Scandinavian countries have also become more ethnically heterogenous due to immigration.

[16] The notion that members of small ethnic groups identify ethnically is consistent with the empirical evidence in Scheve and Slaughter (2001) showing that immigrants have more favorable attitudes toward immigration, also when income is controlled for.

[17] In all cases, it is a poor argument for not allowing the citizens of Singapore to vote.

The idea that ethnic identification is stronger the smaller is the ethnic group is also in line with the study of ethnic minorities in the UK by Bisin et al. (2006). They find evidence that the higher is the percentage of a person's own ethnic group in the neighborhood, the lower is the degree of ethnic identification and the probability of homogamy. Similarly, Fryer and Torelli (2006) find that the phenomenon of 'acting white' among blacks – interpreted as racial differences in the relationship between academic performance and popularity – is stronger in US schools with few black students.

However, it is also plausible that the relationship between ethnic identification and the size of ethnic groups might be different for very small ethnic groups in ways not captured by our model. For example, Miguel and Posner (2006) find that there is a *negative* relationship between the degree of ethnic identification and ethnic diversity for twelve ethnically diverse African countries.

Proposition 2 requires quite strong assumptions on the distribution of income and size of ethnic groups. We therefore specify a simpler version of the model with only two ethnic groups, which allows us to study the effects of ethnic diversity and income inequality under less restrictive assumptions.

3. Black and White Model

In the remainder of the paper, we consider a simpler model with two social classes, poor (P) and rich (R), and two ethnic groups, black (B) and white (W). This simplification is relevant for the US, where the main ethnic division has traditionally been between the Afro-American population and people of European origin. Based on survey data on self-reported social class, the restriction to two social classes seems relevant for the US.[18] The simplified model is also likely to be relevant for other countries – for example ethnic divisions between the native population and non-European immigrants in Europe and between the French and English speaking population of Canada.

We denote the proportion of the four different types in the population by p_{PB}, p_{RB}, p_{PW} and p_{RW} and as before we assume that all four types are represented in the population. In addition, we assume that no type, or sum of two or three types, consists of exactly half the population since this allows us to disregard the possibility of the median falling between two types' preferred tax rate.

All individuals of a certain type are identical – all people in the rich income group have pre-tax income y_R and everybody in the poor group have income y_P satisfying

[18] For example, in the General Social Surveys 1972-2004 (Davis et al. 2005), 46 percent of respondents classify themselves as working class, whereas 46 percent classify themselves as middle class. Of the remaining 8 percent, 5 percent classify themselves as lower class and 3 percent as upper class.

$y_R > y_P > 0$.[19] This specification implies that the status of the ethnic groups is in between the status of the poor and rich groups. In other words, the ethnic identity is the high status identity for poor people, whereas it is the low status identity for rich people.

Actual income distributions are typically skewed so that the median income is less than the average income. Since there are only two income levels in the model, we therefore assume that the poor population is in majority, i.e., $p_{PB} + p_{PW} > 1/2$. Without loss of generality, we also assume that the white population is in majority, i.e., $p_{PW} + p_{RW} > 1/2$. Given these assumptions, we have two different cases. First, if poor whites are in majority, the tax rate is uniquely determined by their identity choice. Second, if poor whites do not constitute a majority of the population, both poor whites and poor blacks could potentially determine the tax rate. The propositions below focus on the second case which is more complicated, while we discuss the outcome when poor whites are in majority informally. We make two different assumptions about the relative income of whites and blacks – we either assume that the white and black population have the same average income, i.e., $p_{RW}/p_{PW} = p_{RB}/p_{PB}$ or that the white population is on average richer, i.e., $p_{RW}/p_{PW} > p_{RB}/p_{PB}$.

The cognitive distance function $d(\cdot)$ is given by the following table:

	Poor black	Poor white	Rich black	Rich white
Black	$\alpha\frac{p_{RB}}{p_{PB}+p_{RB}}$		$\alpha\frac{p_{PB}}{p_{PB}+p_{RB}}$	
White		$\alpha\frac{p_{RW}}{p_{PW}+p_{RW}}$		$\alpha\frac{p_{PW}}{p_{PW}+p_{RW}}$
Poor	$\beta\frac{p_{PW}}{p_{PB}+p_{PW}}$	$\beta\frac{p_{PB}}{p_{PB}+p_{PW}}$		
Rich			$\beta\frac{p_{RW}}{p_{RB}+p_{RW}}$	$\beta\frac{p_{RB}}{p_{RB}+p_{RW}}$

This linear specification implies that it is costless to identify with a group where everybody is of the same type as oneself, whereas the cost goes to α or β when there are nobody like oneself in that group.

We now turn to determining the equilibria of this model. First, recall from (2.2) that the optimal tax rate of someone belonging to social class j and ethnic group k that identifies with his social class j is given by:

$$t^*_{jkj} = \max\left\{1 - \frac{y_j}{y}, 0\right\}. \tag{3.1}$$

It is clear that rich people who identify themselves as rich prefer a zero tax rate (since $y_R > y$) and that poor people who identify themselves as poor prefer the tax rate

[19] It would admittedly be more realistic to allow agents with different individual incomes, but that makes it difficult to derive interesting comparative statics results. In the Appendix we nevertheless consider this alternative and show that the result in Proposition 3 is similar in that setting.

$1 - y_P/y$. Similarly, someone identifying with her ethnic group k prefers the tax rate

$$t^*_{jkk} = \max\left\{1 - \frac{y_j + \gamma y_k}{(1+\gamma)\,y}, 0\right\}. \tag{3.2}$$

The optimal voting choices (3.1) and (3.2) imply that preferred tax rates can be ordered within ethnic groups – for example, rich whites always prefer lower taxes than poor whites. Since we need these results later, we state them formally.

LEMMA 1. *Optimal tax rates always satisfy the following:*
(i) $0 = t^*_{RWR} \leq t^*_{RWW} \leq t^*_{PWW} \leq t^*_{PWP} = 1 - y_P/y$,
(ii) $0 = t^*_{RBR} \leq t^*_{RBB} \leq t^*_{PBB} \leq t^*_{PBP} = 1 - y_P/y$.
In addition, if whites are richer than blacks, then $t^*_{RWW} \leq t^*_{RBB}$ *and* $t^*_{PWW} \leq t^*_{PBB}$.

PROOF. The result follows directly from (3.1) and (3.2) once it is noted that $y_P < y_W < y_R$ and $y_P < y_B < y_R$ for the first part, and $y_W > y_B$ for the second. □

It should be noted that we cannot say whether the preferred tax rate of poor whites identifying themselves as white (t^*_{PWW}) is higher or lower than the tax preferred by rich blacks identifying themselves as black (t^*_{RBB}), even if we make the additional assumption that whites are richer on average. The reason is that the relation between these two tax rates also depends on the parameter γ, i.e., how much individuals care about the status of the group they identify with.

Lemma 1 also implies that when whites are richer than blacks and all types identify with their ethnic group, blacks prefer the same or higher taxes than whites holding income constant. Fong (2001), Alesina et al. (2001) and Alesina and La Ferrara (2005*b*) show empirically that white people in the US are more negative toward redistribution than Afro-Americans also when personal income is held constant. This suggests that the poor in the US identify along ethnic lines rather than with their social class.

Let us now turn to identity choices. The condition (2.3) for high status identification can be rewritten as follows for poor blacks and poor whites:

$$PB : (1-t)\frac{p_{RB}}{p_{PB}+p_{RB}}(y_R - y_P) > \frac{\delta}{\gamma}\left(\alpha\frac{p_{RB}}{p_{PB}+p_{RB}} - \beta\frac{p_{PW}}{p_{PB}+p_{PW}}\right), \tag{3.3}$$

$$PW : (1-t)\frac{p_{RW}}{p_{PW}+p_{RW}}(y_R - y_P) > \frac{\delta}{\gamma}\left(\alpha\frac{p_{RW}}{p_{PW}+p_{RW}} - \beta\frac{p_{PB}}{p_{PB}+p_{PW}}\right). \tag{3.4}$$

When choosing an identity, a poor agent trades off the relatively higher status of the ethnic identity, i.e., the left hand side of (3.3) and (3.4), against the potentially shorter distance to the class identity, i.e., the right hand sides of the two inequalities. For example, the lower the tax rate, the higher is the relative status of the ethnic identity and therefore the poor are more prone to identify with their ethnic groups. Similarly,

the poor are more likely to identify with their ethnic groups the more important is status.

If whites are richer than blacks, the status of the ethnic identity is higher for poor whites than for poor blacks. However, that whites are richer also means that the distance for poor whites to the white identity is larger than the distance for poor blacks to the black identity. It is possible to show that the latter effect dominates and that poor blacks always identify as black if poor whites identify as white. This result is dependent on the linear specification of the distance function and the assumption that the class and ethnic tension parameters are the same for all types, but it is also plausible – if the poor in the majority group favor their ethnic group, then we would probably not expect the poor in the minority group to identify with the poor.

LEMMA 2. *If whites have the same or higher average income than blacks and poor whites identify as white, then poor blacks identify as black.*

PROOF. In the Appendix. □

From Proposition 1 we know that at least one equilibrium exists. Since we have assumed that the poor are in majority, we can show that only the identity choices of the poor matter for the equilibrium tax rate. When whites and blacks are equally rich, there can only be two different tax rates in equilibrium since poor whites and poor blacks prefer the same tax rate when they make the same identity choice. For simplicity, we denote the two possible equilibrium tax rates the poor ($t^*_{PWP} = t^*_{PBP}$) and ethnic ($t^*_{PWW} = t^*_{PBB}$) tax rate.

LEMMA 3. *If blacks and whites have the same average income and poor whites are in minority, then there are two feasible equilibrium tax rates:*

(1) *If poor whites identify as white or poor blacks identify as black, the equilibrium tax rate will be the ethnic tax rate* ($t^*_{PWW} = t^*_{PBB} = 1 - (y_P + \gamma y) / (1 + \gamma) y$).
(2) *If poor whites and poor blacks identify as poor, the equilibrium tax rate will be the poor tax rate* ($t^*_{PBP} = t^*_{PWP} = 1 - y_P/y$).

PROOF. In the Appendix. □

If white people are richer than black people, poor whites and poor blacks prefer different tax rates when they identify with their ethnic group. There can thus be three different tax rates in equilibrium. We refer to these as the poor ($t^*_{PWP} = t^*_{PBP}$), black (t^*_{PBB}) and white (t^*_{PWW}) tax rates.

LEMMA 4. *If whites are on average richer than blacks and poor whites are in minority, then there are three feasible median tax rates in equilibrium:*

(1) *If poor whites identify as white the equilibrium tax rate will be the white tax rate* ($t^*_{PWW} = 1 - (y_P + \gamma y_W) / (1+\gamma) y$).
(2) *If poor whites and poor blacks identify as poor the equilibrium tax rate will be the poor tax rate* ($t^*_{PBP} = t^*_{PWP} = 1 - y_P/y$).
(3) *If poor whites identify as poor, but poor blacks as black, the equilibrium tax rate will be the black tax rate* ($t^*_{PBB} = 1 - (y_P + \gamma y_B) / (1+\gamma) y$).

PROOF. In the Appendix. □

Before we turn to more interesting comparative statics, we discuss how some parameters affect the equilibrium in order to demonstrate the mechanisms in the model. An increase in the weight of status in the utility function (as measured by γ) has two effects. First, ethnic identities become more attractive for the poor, which makes it more likely that they identify themselves with their ethnic group. Second, the tax rates preferred by the poor when they identify with their ethnic identity decrease since they give more weight to their ethnic group (which is richer than the poor) in the voting decision. These two effects go in the same direction so that for sufficiently high γ both poor whites and poor blacks identify with their ethnic group, and the equilibrium tax rate is the white tax rate (which is zero for large enough γ). On the other hand, as γ approaches zero, the income of the ethnic group no longer affects the preferred tax rate of the poor, implying that the voting choices of the poor are independent of their identity choices. The tax rate therefore approaches the poor tax rate, which correspond to the standard model of redistribution without identity choices.

How the weight of cognitive distance, δ, affects identity choices depends on the agents' relative distances to the class and ethnic identities. If the cognitive distances for poor blacks and poor whites to the poor identity are larger than the distances to their ethnic identities, they always identify with their ethnic group. If this is not the case, a higher δ makes it more likely that the poor identify as poor.

It is clear that there will be a unique equilibrium when these or other parameters approach extreme values since either identity becomes superior regardless of tax rates.[20] However, for intermediate parameter values there might be multiple equilibria. As suggested by the proof of Proposition 1 there is a complementarity in the model – a higher tax rate makes it more likely that the poor switch to the poor identity and prefer higher taxes. To illustrate the possibility of multiple equilibria, we consider a simple parametric example in which blacks and whites are equally rich.[21] In this example,

[20] This is true unless the cognitive distances to the class and ethnic identities happen to be exactly the same.

[21] The parameters used in this example are $p_{PW} = 0.45$, $p_{RW} = 0.30$, $p_{PB} = 0.15$, $p_{RB} = 0.10$, $y_P = 100$, $y_R = 300$, $\gamma = 0.5$, $\delta = 20$, $\alpha = 4$ and $\beta = 0.5$.

the rich are three times richer than the poor and there are 25 percent blacks in the population. The status parameter γ is 0.5, implying that an agent gives the after-tax income of the group he identifies with half the weight of his personal after-tax income. The cognitive distances to the poor identity must be smaller than the distances to the ethnic identities in order for the poor to identify as poor. The class tension parameter α is therefore set relatively high compared to the ethnic tension parameter β. The psychological cost of cognitive distance is modest compared to the utility derived from personal income. The utility cost for poor whites in terms of cognitive distance when they identify as white is 32, and 2.5 when they identify as poor. This utility cost can be compared to the utility of 100 that the poor derive from personal income when the tax rate is zero.

In this example, poor whites identify as poor if the tax rate is above 26 percent, whereas poor blacks identify as poor if the tax rate is above 39 percent. From Lemma 3 we know that the poor tax rate, which in this example is 44 percent, prevails when both types identify as poor. When poor blacks identify as black or both poor whites and poor blacks identify with their respective ethnic groups, the tax rate is instead 30 percent. Both of these tax rates can be sustained in equilibrium – in the high tax equilibrium both poor blacks and poor whites identify with their class, whereas the poor blacks identify as black in the low tax equilibrium.

The possibility of multiple equilibria implies that an equilibrium can be suboptimal in the sense that each agent of a certain type would reach a higher utility level if the other agents of the same type changed identity and preferred tax rate. However, given the identity choices of the other agents, no agent has an incentive to change identity or vote differently.[22] For example, we might have one high tax equilibrium where poor whites identify as poor and one low tax equilibrium where they identify as white. Based on Marxian theory it might be tempting to conclude that the poor are better off in a high tax equilibrium, and that the poor should be made "class conscious" if the low tax equilibrium prevails. However, although a class identity would benefit their material interest, our model allows no such conclusion since agents also get utility from their identity – it may well be the case that the poor's utility is lower in a high tax than in a low tax equilibrium.[23]

[22] Since voting and identity choices are made separately, an agent can end up in a suboptimal equilibrium even if he is the only agent in the economy (and cognitive distances are defined so that this is possible). A single agent might prefer to simultaneously switch identity and preferred tax rate, but this is ruled out by the definition of an equilibrium.

[23] More generally, the idea that people may hold dysfunctional identities is often raised in the literature on identity and may be important in order to understand self-destructive behaviors such as "ghetto culture" (see Akerlof and Kranton 2000 for references and further discussion).

Although differences in redistribution can be explained in terms of multiple equilibria for identical parameter values, we now go on to study how the set of potential equilibria changes with the parameters of the model. These results together with some empirical evidence are presented in the following three sections.

4. Ethnic Diversity

The main lesson from Proposition 2 is that ethnic diversity might induce the poor to identify with their ethnic group and therefore favor lower taxes. In this section we will see that this conclusion does not hold universally.

In the black and white model, blacks constitute a minority and we therefore model an increase of ethnic diversity as an increase in the black population. The results differ depending on whether poor whites are in majority or not. When poor whites are in majority, their identity choices alone determine the tax rate. When poor whites are in minority, the identity choices of poor blacks also affect the tax rate (unless poor whites already identify as white). The propositions focus on the comparative statics in the more complicated latter case, but we will also discuss the case when poor whites are in majority.

The effect of ethnic diversity depends both on whether poor whites are in majority and the extent of interethnic income inequality. Proposition 3 focus on an increase in the proportion of blacks when whites and blacks have the same average income. Proposition 4 instead focuses on the case when blacks are poorer. Finally, Proposition 5 studies an increase in the number of poor blacks when blacks are poorer. Table 1 summarizes the overall effect on the level of redistribution in these three different cases. As is clear from Table 1, the effects of ethnic diversity is typically not monotonic.

TABLE 1. Effects on redistribution of an increase in ethnic diversity

	Poor whites are in majority	Poor whites are in minority
Increase of blacks, no interethnic inequality (Proposition 3)	$-$	$+$
Increase of blacks, interethnic inequality (Proposition 4)	$-$	$-/+$
Increase of poor blacks, interethnic inequality (Proposition 5)	$-/+$	$-/+$

The remainder of this section derives and explains the findings reported in Table 1 with special emphasis on explaining the non-monotonicities. The simplest case is to assume that blacks and whites are equally rich and that the proportion of poor and rich blacks increases proportionally, so that the average incomes of the black and white

groups are held constant. In this case, the only effect of an increase in the proportion of blacks on identity choices is to increase the cognitive distance to the class identity for poor whites and decrease it for poor blacks. The relative status of both identities and distance to the ethnic identity is unaffected by changes in ethnic diversity. As the proportion of blacks increases, poor blacks therefore become more prone to identify as poor whereas poor whites become more prone to identify as white. In contrast to Proposition 2, the tax rate cannot decrease as the proportion of blacks increases.[24] The reason is that if poor whites switch to the white identity, poor blacks already identify as black (recall Lemma 2) and the ethnic tax rate prevails.[25]

PROPOSITION 3. *If poor whites are in minority and blacks and whites have the same average income, then an increase in the black population, while keeping the average income of blacks and whites constant, implies the following for the equilibrium tax rate:*

(1) *If poor whites initially identify as white or both poor whites and poor blacks identify as poor, then the tax rate is unchanged.*
(2) *If poor whites initially identify as poor and poor blacks identify as black, then poor blacks might switch to the poor identity resulting in a higher equilibrium tax rate.*

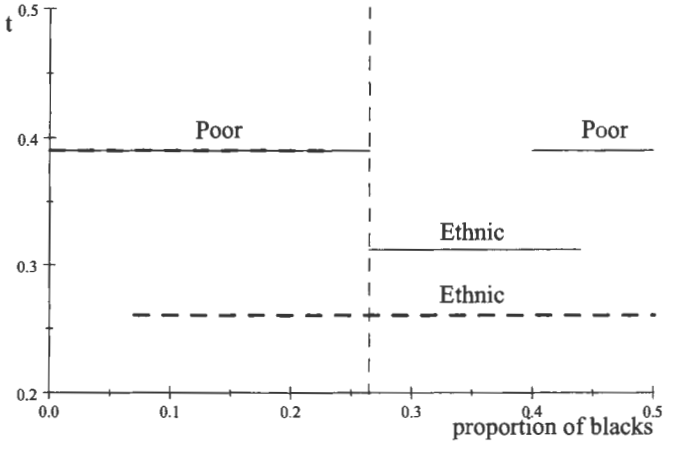
PROOF. In the Appendix. □

Proposition 3 only refers to the case when poor whites are in minority. Hence it doesn't apply when the proportion of blacks is so small that poor whites are in majority. To describe the full comparative statics, i.e., when p_B goes from zero to infinitely close to one half, we consider two different parametric examples illustrated in Figure 1. The thin dashed vertical line in Figure 1 indicates the proportion of blacks above which poor whites are in minority.

[24] Note that Proposition 3 isn't a special case of Proposition 2 since the black group doesn't meet the size restriction in Proposition 2 when they initially identify as black.

[25] In the Appendix, we show that the tax rate may decrease when the proportion of blacks increases if there is heterogeneity in incomes within social classes. The reason is that the tax rate when poor whites identify as white is lower than the tax rate prevailing when poor blacks identify as black and poor whites as poor.

FIGURE 1. Increase in ethnic diversity (no interethnic income inequality)



Thick dashed lines: High status ($\gamma = 0.5$). Thin lines: Low status ($\gamma = 0.25$).

The thick dashed line in Figure 1 indicates the equilibrium tax rate in the first parametric example as a function of the proportion of blacks.[26] In this first example, agents care relatively much about the status of the group they identify with (the status parameter γ is 0.5). Recall that the more important status is, the more likely it is that poor blacks and poor whites will identify with their respective ethnic groups. In this case, poor blacks always identify themselves as black. Poor whites, on the other hand, identify as poor when society is ethnically homogenous (less than 7 percent blacks) and the higher poor tax rate is the only possible equilibrium. The poor tax rate is an equilibrium also when the proportion of blacks is between 7 and 24 percent. However, since poor whites now identify as white at the lower ethnic tax rate, this can also be an equilibrium. When the proportion of blacks is above 24 percent, poor whites identify as white at all tax rates and only the ethnic tax rate is an equilibrium. Hence, poor whites already identify as white at the point when they become a minority (at 27 percent blacks), implying that the tax rate is unaffected by this shift in potential majorities.

Note the contrast between this example and Proposition 3. In this example, the tax rate is roughly monotonically decreasing in ethnic diversity, whereas Proposition

[26] The parameters used in this example are $p_P = 0.68$, $y_P = 100$, $y_R = 300$, $\gamma = 0.50$, $\delta = 20$, $\alpha = 4$ and $\beta = 1.3$.

3 implies that ethnic diversity might lead to higher tax rates. The reason is that Proposition 3 focuses on the case when poor whites are in minority and implicitly rules out the effect of poor whites switching from the poor to the white identity. At the level of ethnic diversity where the poor whites switch to the white identity, the low tax equilibrium already prevails since poor blacks already identify as black (recall that Lemma 2 implies that poor blacks always identify as black if poor whites identify as white).

The equilibrium tax rate in next parametric example is indicated by the thin lines in Figure 1. This example illustrates that when status is sufficiently unimportant, ethnic diversity need not have a monotonic effect on the equilibrium tax rate. The only difference compared to the previous example is that status is less important ($\gamma = 0.25$ compared to $\gamma = 0.50$), which has two different effects: It makes it more likely that the poor identify with their social class, and it leads to a higher ethnic tax rate. In this example, poor whites always identify as poor. When the proportion of blacks is below 27 percent, poor whites are in majority and since they always identify as poor, the poor tax is implemented. If the proportion of blacks is between 27 and 40 percent, poor whites are not in majority and poor blacks identify as black at both tax rates, so only the ethnic tax rate is an equilibrium. When the proportion of blacks is between 40 and 44 percent, poor blacks identify as poor at the poor tax rate and as black at the ethnic tax rate, implying that both tax rates are equilibria. Finally, as the proportion of blacks is above 44 percent, poor blacks identify as poor at all tax rates and only the poor tax rate can be an equilibrium.

Note that though the effect of ethnic diversity on redistribution was monotonic and negative in the first parametric example, this is not the case in the second example. Instead, redistribution is high when ethnic diversity is either very low or very high. In the intermediate case, there are enough blacks to influence the tax rate, but so few that poor blacks are reluctant to identify with the poor. This provides an explanation for the finding in Dincer and Lambert (2006) that there is a U-shaped relationship between redistribution and ethnic and religious fractionalization.[27]

Proposition 3 only applies to the case when blacks and whites have the same income. In many societies, the minority population is poorer than the majority group. In this case, increasing the size of the minority group decreases the average income in the population, leading to lower tax rates for given identity choices. This is a result of the fact that there are only two income groups in the population – with continuous incomes both the median and average income would typically decrease and the overall effect be

[27] Dincer and Lambert (2006) also study the effects of ethnic and religious polarization on redistribution and find a monotonic relationship. Since we only have two ethnic groups in the model, we cannot distinguish between ethnic fractionalization and polarization.

ambiguous. Proposition 4 focuses on the case when blacks are on average richer, but the number of blacks increases proportionally so that the average income of both blacks and whites are held constant. In this case, the result is very similar to Proposition 3. One difference, however, is that there are three instead of two potential equilibrium tax rates. Since the black group is poorer than the white group, poor blacks now favor higher taxes when they identify as black than poor whites when they identify as white. The case when the poor whites are in minority is described in Proposition 4.

PROPOSITION 4. *If poor whites are in minority and whites are on average richer than blacks, then an increase in the black population, while keeping the average income of blacks and whites constant, implies the following for the equilibrium tax rate:*

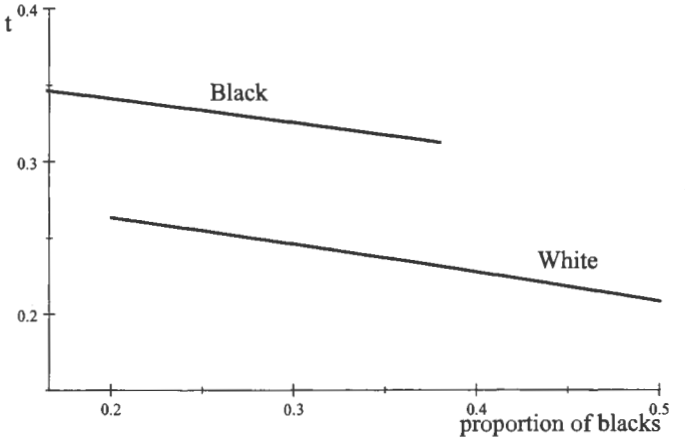
(1) *If poor whites initially identify as white or both poor whites and poor blacks identify as poor, the equilibrium tax rate decreases.*
(2) *If poor whites initially identify as poor and poor blacks identify as black, then either poor blacks or poor whites may switch identity and the equilibrium tax rate may therefore either increase or decrease.*

PROOF. In the Appendix. □

Proposition 4 is illustrated by a parametric example in Figure 2 where we only consider the case when poor whites are in minority.[28] One difference compared to the previous parametric examples is that since the average income decreases in ethnic diversity, tax rates for given identity choices are decreasing in the proportion of blacks. In this example, poor blacks identify as black irrespective of the tax rate. Poor whites identify as poor when the proportion of blacks is below 20 percent. If there are sufficiently many blacks for poor whites to be in minority (17 percent), the equilibrium tax rate is therefore the black tax rate. When the proportion of blacks is between 20 and 38 percent, poor whites still identify as poor at the black tax rate, but as white at the white tax rate, implying that we have multiple equilibria. As the tax rate increases above 38 percent, poor whites identify as white also at the black tax rate, and so only the white tax rate is an equilibrium.

[28] The parameters used in this example are $p_{P|B} = 0.8$, $p_{P|W} = 0.6$, $y_P = 100$, $y_R = 300$, $\gamma = 0.5$, $\delta = 20$, $\alpha = 4$ and $\beta = 0.5$. These parameters imply that the average income of the white group is 29 percent higher than the average income of the black group, which is lower than the actual income difference between blacks and whites in the US.

FIGURE 2. Increase in ethnic diversity (interethnic income inequality)



In both Proposition 3 and 4, the number of poor blacks and rich blacks increases proportionally, ensuring that the incomes of the black and white groups are held constant. In many cases, for example immigration, it is more reasonable to assume that it is only the proportion of poor blacks that increases. This implies a change in income inequality both within and between the two ethnic groups, which introduces a more intricate interaction between social class and ethnicity.

Proposition 5 establishes what can be said generally when the proportion of blacks among the poor increases and the average income in the population is held constant.[29] As the number of blacks among the poor increases, the white group becomes richer and the status of the white identity increases. Since there are more poor blacks in the population, the distance to the poor identity increases for poor whites. These two effects tend to push poor whites toward the white identity, but are counteracted by the increase in the cognitive distance to the white group (since the proportion of poor whites among whites goes down). As the proportion of poor whites among whites decreases at a faster rate the lower is the initial proportion of poor whites, the latter effect dominates if there are sufficiently many blacks among the poor. The exact threshold level ($\bar{p}_{B|P}$) is specific to our linear specification of the cognitive distance

[29] The main reason for holding the average income constant is that the comparative statics become considerably more complicated if also the average income changes. In addition, decreasing the average income have an unambigous negative effect on the tax level, but this result is an artefact of the fact that their are only two income groups (otherwise both the median and average would change).

function, but, as can be seen from Proposition 5, it is increasing in the proportion of rich whites to the poor.

Similarly, the effect of more blacks among the poor on the identity of poor blacks is not guaranteed to go in the direction of class identification. On the one hand, both the status of the black group and the distance to the poor identity decreases, but on the other hand the distance to the black identity decreases since the black group now consists of relatively more poor blacks.[30] The latter effect dominates as long as the proportion of blacks among the poor is below a certain threshold ($\hat{p}_{B|P}$). This suggests that if the size of a small minority group is increased with poor people, it might push the poor of the minority group toward ethnic identification.

PROPOSITION 5. *If poor whites are in minority and whites are on average richer than blacks, then an increase of blacks among the poor ($p_{B|P}$), while the average population income is held constant, implies the following for the equilibrium tax rate:*

(1) *If poor whites initially identify as white and the proportion of blacks among the poor is sufficiently low ($p_{B|P} < \bar{p}_{B|P}$), then the equilibrium tax rate decreases and identity choices of poor whites are unchanged.*

(2) *If both poor whites and poor blacks initially identify as poor and the proportion of blacks among the poor is sufficiently low ($p_{B|P} < \hat{p}_{B|P}$), then poor blacks (and possibly also poor whites) may switch to the ethnic identity which decreases the tax rate.*

(3) *If poor whites identify as poor and poor blacks as black and the proportion of blacks among the poor is sufficiently high ($p_{B|P} > \bar{p}_{B|P}$), then the equilibrium tax rate will increase and poor blacks may switch to the poor identity which increases the tax rate further.*

The threshold $\bar{p}_{B|P}$ is given by $1/2+p_{RW}/2\,(p_{PB}+p_{PW})$ and the threshold $\hat{p}_{B|P}$ is given by $1/2-p_{RB}/\,(p_{PB}+p_{PW})$.

PROOF. In the Appendix. □

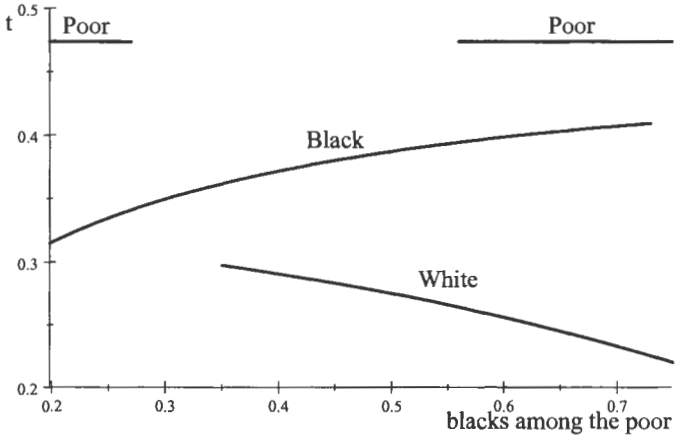
In order to demonstrate some of the different effects summarized in Proposition 5, we again use a simple parametric example.[31] Figure 3 only shows the case when poor whites are in minority. As can be seen from Figure 3, the black tax rate is increasing and the white tax rates is decreasing in the proportion of blacks among the poor. The reason is that the average income of the black group is decreases, while the average income of the white group increases.

[30] Similarly, the status gain of choosing an ethnic identity goes down, and the sign of the total effect on the identity choice depends on the relative importance of status and class hostility.

[31] The parameters used in this example are $p_{B|R}=0.2$, $p_P=0.55$, $y_P=100$, $y_R=300$, $\gamma=0.5$, $\delta=20$, $\alpha=4$ and $\beta=0.7$.

First, consider the case when there are 20 to 27 percent blacks among the poor. In this case, the poor whites identify as poor at all tax rates, whereas poor blacks only identify as poor if the highest tax rate prevails. This implies that the highest tax rate is an equilibrium. However, since poor blacks identify as black at the black tax rate, this can also be an equilibrium. As the number of blacks among the poor increases slightly, but not above 35 percent, poor blacks identify as black and poor whites as poor at all tax three tax rates, so the black tax rate is the only possible equilibrium. This demonstrates that an increase of low-skilled workers from a minority group might push these workers toward ethnic identification and lower tax rates. Hence, whereas Proposition 3 and 4 showed that increasing ethnic diversity may induce the poor from a minority group to identify with their social class instead of their ethnic group, this result might be overturned when the minority group also becomes poorer. In other words, lower redistribution as a result of ethnic diversity might be driven by the demands of the poor minority group.

FIGURE 3. Increase in the proportion of blacks among the poor



When the number of blacks among the poor is between 35 and 56 percent, poor whites identify as white at the white tax rate, but as poor at higher tax rates. Poor blacks identify as black at all tax rates, so both the white and black tax rates are equilibria. In the interval between 56 and 73 percent blacks among the poor, poor whites identify as white at the white tax rate, but as poor at the black and poor tax rate, whereas poor blacks identify as black at the white and black tax rates. This implies

that all three tax rates are equilibria. Finally, when the number of blacks among the poor is above 73 percent, but not so many that blacks are in majority, poor blacks identify as poor at the black tax rate so that there are again two equilibria – either the poor or the white tax rate.

The example in Figure 3 illustrates that identity choices of one group may depend on the identity choices of another group. When there are 56 to 73 percent blacks among the poor, poor blacks identify as black if poor whites identify as white. However, if poor whites instead identify as poor, poor blacks might identify as poor. When there are more than 73 percent blacks among the poor, poor blacks identify as poor for certain if poor whites identify as poor, whereas they identify as black if poor whites identify as white.

In this example, the high tax equilibrium where both poor blacks and poor whites identify as poor is only possible for a very low and a very high degree of ethnic diversity, whereas it is not possible for intermediate levels of diversity (i.e., between 27 and 56 percent of blacks among the poor). This U-shaped relationship between redistribution and ethnic fractionalization is in line with the empirical finding of Dincer and Lambert (2006). Recall that a similar effect was shown in the example in Figure 3 where whites and blacks had the same average income. However, in that case, the non-monotonicity was induced by a shift in the set of potential majorities, whereas in this case it is the actual identity choices that are non-monotonic. Proposition 5 thus provides a caveat to the conjecture in Myrdal (1944) and Lipset and Marks (2006) that the high degree of ethnic fragmentation among the working class was one factor that prevented the development of a strong workers' movement in the United States. As demonstrated by Figure 3, this conclusion may not hold universally.

Propositions 3 to 5 may shed some light on the evidence on class voting, i.e., the extent to which social class determines voting behavior (see Nieuwbeerta and Ultee 1999 for references to this sociological literature). The model suggests that immigration of foreign low-skilled people might induce poor whites – and possibly also poor blacks – to identify with their ethnic group and support lower taxes. The inflow of relatively poor immigrants may therefore be part of the explanation for why class voting has declined in Europe during the last decades, as well as why European anti-immigration political parties seem to have gained in popularity.[32] The latter is supported by empirical studies by Knigge (1998) and Golder (2003) showing that the support for anti-immigration parties is indeed increasing in the level of immigration. A competing explanation for the relatively strong support that anti-immigration parties get from the working

[32] Examples of such parties include FPÖ (Austria), Schweizerische Volkspartei (Switzerland), Dansk Folkeparti (Denmark), Vlaams Blok (Belgium), Fremskridtspartiet (Norway) and Front national (France).

class is a fear for increased competition in the labor market. However, in contrast to our model, this does not explain why these parties often advocate a low level of redistribution (see for example Betz 1993, Poglia Mileti et al. 2002 and McGann and Kitschelt 2005).[33] In addition, the empirical evidence on the relationship between support for anti-immigration parties and the level of unemployment is ambiguous (see Knigge 1998 and Golder 2003).[34]

Proposition 5 shows the result of simultaneously increasing ethnic diversity and interethnic income inequality. In the next section we turn to studying the effects of income inequality – both between classes and ethnic groups – separately from the effects of ethnic diversity.

5. Income Inequality

Income inequality can mean two different things in this model – the income difference between social classes and the difference in income between ethnic groups. We first analyze the effects of income inequality between rich and poor.

Standard models of income redistribution, e.g. Meltzer and Richard (1981), imply that redistribution increases as a response to an increase in pre-tax income inequality as measured by the distance between the average and median income. When income inequality increases, the poor become poorer compared to the rich which increases their demand for redistribution. In our model there is a counteracting effect since an increase in income inequality increases the status of ethnic identities, which might lead to a shift to ethnic identities and lower tax rates. Since the comparative statics are considerably more complicated if blacks are poorer than whites, without bringing many additional insights, Proposition 6 is only stated for the case when whites and blacks have the same average income.

PROPOSITION 6. *If poor whites are in minority and if whites and blacks have the same average income, then an increase in pre-tax income inequality* $(y_R - y_P)$*, while average income* y *is held constant, implies the following for the equilibrium tax rate:*

(1) *If poor blacks identify as black, the tax rate increases. Furthermore, if in addition income inequality is high* $(y_P/y < (1-\gamma)/2)$*, poor blacks (and possibly*

[33] The political bundling effect in a two-dimensional policy space demonstrated by Roemer (1998) can explain why a voter could vote for a right-wing party although she favors a high degree of redistribution, but not why anti-immigration parties tend to focus on right-wing economic policies in the first place.

[34] Of course, the absence of any relation between the unemployment rate and anti-immigration sentiments is only an argument against the labor market hypothesis if agents are not perfectly forward-looking regarding the effects of increased immigration, but adjust their beliefs about negative effects of immigration in response to a high level of unemployment.

poor whites) might switch to the poor identity which increases the tax rate further. If income inequality instead is low ($y_P/y > (1-\gamma)/2$), *the identity choices of poor blacks are unchanged.*

(2) *If poor whites and poor blacks initially identify as poor, the tax rate increases and the identity choices of the poor are unchanged if income inequality is high* ($y_P/y < 1/2$). *If instead income inequality is low* ($y_P/y > 1/2$), *poor blacks (and possibly poor whites) might switch to ethnic identities which might lead to a lower tax rate.*

PROOF. In the Appendix. □

As can be seen in Proposition 6, the effect of an increase in pre-tax income inequality depends on the initial degree of income inequality. If income inequality is initially high, the tax rate increases so much in response to an increase in pre-tax income inequality that after-tax income inequality decreases, which decreases the relative status of the ethnic identities. On the other hand, if income inequality is initially low, after-tax income inequality increases and the ethnic identities becomes more attractive.

To see why this is the case, note that an increase in income inequality has two counteracting effects on the relative status of ethnic and poor identities. For given tax rates, higher pre-tax income inequality implies that the ethnic identities become more attractive for poor blacks and poor whites. On the other hand, for given identity choices, the tax rate increases as a response to higher inequality, which makes the poor identities more attractive. To see these two effects, recall from the conditions for ethnic identification (3.3) and (3.4) for poor blacks and poor whites that the status difference between the ethnic and class identities is some population parameter times $(1-t)(y_R - y_P)$. Differentiating with respect to $y_R - y_P$ gives

$$\frac{\partial (1-t)(y_R - y_P)}{\partial (y_R - y_P)} = (1-t) - \frac{\partial t}{\partial (y_R - y_P)}(y_R - y_P). \tag{5.1}$$

The first term in this expression is the direct effect of income inequality, whereas the second term is the effect through the increase in the tax rate. Since $\partial t/\partial (y_R - y_P)$ does not depend on $y_R - y_P$, we see that this latter effect is stronger if income inequality is initially high.

The result that an increase in income inequality has ambiguous effects on redistribution fits well with recent empirical evidence showing no clear connection between income inequality and redistribution (see Perotti 1996 and Lind 2005). However, our model is not the first to produce this result. For example, in Corneo and Grüner (2000) the median voter prefers less redistribution as economic inequality increases, since the cost of taxation in terms of lost social prestige relative to the working class increases

with economic inequality. The result that an increase in pre-tax income inequality might induce the poor to switch identity and thus favor lower taxes is also present in a slightly different flavor in Shayo (2005).

Although the effect of pre-tax income inequality on redistribution is ambiguous, Lind (2007) argues that interethnic inequality reduces the support for redistribution.[35] To study this effect in our model, we model income inequality between blacks and whites as an increase in the number of poor blacks and a corresponding decrease in the number of poor whites, while the total number of poor and blacks is held constant.[36] This change clearly decreases the status of the black identity, whereas the status of the white identity increases. Cognitive distances to both ethnic and class identities are affected. For poor blacks, the distances to the black and poor identities shrink since there are more poor blacks among the black, but also more poor blacks among the poor. For the same reasons the opposite holds for poor whites – the distances to the white and poor identities increase. In addition, ethnic tax rates are also affected since blacks become poorer and whites become richer. Proposition 7 establishes the different cases when the net effect on the tax rate can be established.

PROPOSITION 7. *If poor whites are in minority and blacks are on average poorer than whites, then an increase in the proportion of poor among blacks, holding average income and the proportion of blacks constant, implies the following for the equilibrium tax rate:*

(1) *If poor whites identify as white and there are fewer blacks than rich people, then the tax rate decreases (and identity choices of poor whites are unaffected).*
(2) *If poor whites and poor blacks identify as poor, then poor blacks (and possibly poor whites if there are fewer blacks than rich people) may switch to ethnic identities which decreases the tax rate.*
(3) *If poor whites identify as poor, poor blacks as black and there are more blacks than rich people, then the tax rate increases and poor blacks might switch to the poor identity which increases the tax rate further.*

PROOF. In the Appendix. □

As shown by the proof of Proposition 7, poor blacks become more likely to identify as black when the number of poor among blacks increases (for a given tax rate). This

[35] Lind (2007) shows this theoretically in a model where people's altruism are targeted towards their own group. He also provides somewhat weak empirial support that between group inequality reduces the support for redistribution (using U.S. panel data from 1969 to 2000).

[36] Studying interethnic inequality in this way implies that the cognitive distances are affected, whereas these are unaffected by a change in standard income inequality. In a model with more than two income groups, interethnic inequality could instead be analyzed as income changes that wouldn't affect cognitive distances.

is a result of the linear specification of cognitive distances – the shorter distance to the black identity turns out to always dominate the effect of lower status of the black identity and shorter distance to the poor identity. For poor whites, the net effect depends on the proportion of blacks to the proportion of rich. The status of the white identity increases, but the distances to both the poor identity and to the white identity increase. If there are fewer blacks than rich in the population, it is certain that poor whites become more likely to identify as white.

Comparing Proposition 6 and 7, it is clear that the effects of standard versus interethnic income inequality typically have opposite effects if income inequality is initially high and the proportion of blacks is lower than the proportion of rich people. This is in line with the theoretical and empirical results in Lind (2007). The novel idea behind our result is that higher income inequality between ethnic groups might induce the poor of the majority group to switch to their ethnic identity in order to enjoy the higher status of the ethnic group.

6. Ethnic and Class Hostility

Apart from the size and economic position of minorities, there are probably also differences in hostility between ethnic groups across countries. For example, a country like the US, with its history of slavery and discriminatory laws, arguably has more tense ethnic relationships than, say, Sweden. In terms of the model, ethnic tension imply that β is high and distances to class identities are large. In turn, it is more likely that individuals identify with their ethnic group, which implies lower taxes and less redistribution.

The model also allows us to study exogenous variation in hostility between social classes. If there are sufficiently weak tensions between rich and poor, i.e., if α is low, both poor whites and poor blacks identify with their ethnic group and in equilibrium the tax is low. Stronger class tension increases the distances to ethnic identities and for sufficiently high α we can be sure that poor blacks and poor whites identify as poor and the tax rate is the highest possible.

Although class hostility as captured by α might vary between countries, beliefs about the causes of poverty may also be important.[37] Alesina et al. (2001) and Fong (2001) show empirically that the belief that poverty is caused by laziness and not bad luck is a strong predictor of negative attitudes toward redistribution.[38] Arguably, if

[37] See for example Piketty (1995), Alesina and Angeletos (2005) and Bénabou and Tirole (2006) for models where such beliefs are endogenously determined.

[38] Gilens (1999, p. 172–173) develops a similar argument: "The belief that black Americans lack commitment to the work ethic is central to whites' opposition to welfare. But it appears that this

the poor believe that the rich have worked hard for their higher incomes, they are less likely to feel aversion toward the rich. Conversely, the rich would feel more sympathetic toward the poor if they thought that poverty was caused by bad luck instead of laziness. Therefore, such beliefs are not captured very well by the class conflict parameter α. However, if we reinterpret α as the strength of the belief that poverty is caused by laziness, a high α implies that the rich feel more distant to their ethnic identity and that the poor closer to their ethnic identity. To incorporate this in the model, we can replace α by $1/\alpha$ in the distance functions to the ethnic identity for poor blacks and poor whites. In this case, a high α tend to push the rich toward class identification, whereas the poor are pushed toward ethnic identification. This provides a simple argument for why beliefs about the causes of poverty may matter for redistribution. Strong beliefs that poverty is caused by laziness make it more likely that the poor identify with their ethnic group, which in turn imply low taxes and little redistribution (compared to the case with class identification). Such beliefs are of course likely to directly affect preferences for redistribution, but the possibility of identity shifts demonstrates an extra channel through which those beliefs may lead to lower redistribution.

7. American Exceptionalism

Why is redistribution so much lower in the US compared to Western Europe? In terms of our model, the US population with European origin is represented by whites, whereas the Afro-American population is represented by blacks.[39] In Western Europe, white refers to the native population and black refers to non-European immigrants.

First, pre-tax income inequality is higher in the US than in Western Europe. On the one hand, the tax preferred by the poor is increasing in income inequality for given identities. On the other hand, if income inequality in the US and Western Europe is lower than the threshold in Proposition 6, then the higher level of income inequality in the US will make the poor more likely to identify with their ethnic group. Hence, the effect on redistribution from the higher pre-tax income inequality in the US is ambiguous.

Second, the higher degree of ethnic diversity in the US may imply that poor whites in the US are more likely to identify as white and favor a low level of redistribution

race-based opposition does not primarily reflect either a general racial animosity or an effort to defend whites' concrete group interests. Rather, the racial component of white opposition to welfare seems to reflect the most important nonracial basis of welfare opposition: the perception that welfare recipients are undeserving."

[39] There are of course many other ethnic groups in the US, but we focus on blacks and whites. This has also been the focus in the literature on racial issues in the US, with Myrdal (1944) as the classic reference. Loury (2002) provides a more recent account on racial stigmatization in the US.

(under the conditions given in Proposition 5). Similarly, to the extent that interethnic income inequality is higher in the US, Proposition 7 shows that this might be an additional force in the same direction. Moreover, the preferred tax rate of poor whites when they identify themselves as white is decreasing in the affluence of whites. Hence, to the extent that whites in the US are more wealthy than their counterparts in Europe, poor whites in the US who identify as white favor a lower tax rate than poor whites in Europe identifying as white, holding everything else constant.

Third, Americans are much more prone than Europeans to believe that the poor are lazy rather than unlucky. In addition, the US has historically a more troubled racial relationship than most European countries. Both of these differences suggest that the poor whites should be more likely to identify as white in the US.

These different explanations do not yield an unambiguous prediction, but, given the argument above, it indeed seems plausible that poor white Americans should be more likely to identify as white and favor low taxes than poor white Europeans. It should be noted, however, that the difference in redistribution between the US and Western Europe could also be rationalized in terms of multiple equilibria.[40] Even if the US and Western Europe were identical in terms fundamentals (i.e., parameter values), it could be the case that poor whites in Europe identify as poor simply because redistribution is relatively high – if redistribution had been at US levels they would have switched to ethnic identities and supported lower taxes.

We are not the first to raise the argument that ethnic diversity is important in explaining differences in redistribution between the US and Europe. Shayo (2005) argues that a high degree of ethnic diversity concentrated to the poorer segments of society induces the poor to identify with their nation instead of their class, thereby reducing the support for redistribution. Alesina et al. (2001) claim that differences in beliefs about the poor and ethnic heterogeneity explains the comparably low level of redistribution in the US through its impact on altruism. However, since altruism is itself an exogenous parameter in their theoretical framework, they do not explicitly model how these factors explain altruism. Moreover, Alesina et al. (2001) consider altruism to be nondiscriminatory across groups, whereas altruism in our model is only directed at a particular subgroup of the population. Lind (2007) studies such directed altruism, but unlike our approach the decision to sympathize with a particular subgroup is not endogenous in his model.

[40] Several other economists have also argued that differences in the level of redistribution across countries can be understood in terms of multiple equilibria. See Alesina and Angeletis (2005) and Bénabou and Tirole (2006) for two recent contributions.

8. Concluding Remarks

The model presented in this paper treats social categories as given despite the fact that such social constructs typically are not unchanged in the longer term (see Alesina and La Ferrara 2005*a* and Posner 2005 for discussion and examples). An interesting and challenging task for future research would therefore be to make social identities, and not only social identification, endogenous. Relatedly, groups may have incentives to manipulate the identity choices of others. For example, the rich might try to reduce the level of redistribution by convincing the poor to identify with their ethnic group. This could be done by directly influencing their identity choice through propaganda, or, more indirectly, by trying to create new ethnic categories.

Appendix A: Proofs

For several of the proofs it is useful to rewrite the two inequalities (3.3) and (3.4) as follows by dividing by $p_{RB}/(p_{PB}+p_{RB})$ and $p_{RW}/(p_{PW}+p_{RW})$, respectively:

$$PB: \frac{p_{PW}}{p_{PB}+p_{PW}}\frac{p_{PB}+p_{RB}}{p_{RB}} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right), \tag{A.1}$$

$$PW: \frac{p_{PB}}{p_{PB}+p_{PW}}\frac{p_{PW}+p_{RW}}{p_{RW}} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right). \tag{A.2}$$

A.1. Proof of Lemma 2. We want show that the left hand side of (A.1) is larger than the left hand side of (A.2). Rearranging this condition we get

$$p_{PW}\frac{p_{RW}}{p_{PW}+p_{RW}} > p_{PB}\frac{p_{RB}}{p_{PB}+p_{RB}}.$$

If white and black have the same average income, we only need to show that $p_{PW} > p_{PB}$. This follows from the fact that whites are in majority and that blacks and whites have the same average income (to see this, divide $p_{RW}+p_{PW} > p_{PB}+p_{RB}$ by p_{PB} and substitute $p_{RB}/p_{PB} = p_{RW}/p_{PW}$). Now suppose that whites are on average richer than blacks. Rewriting the above condition, we get

$$\frac{p_{PW}}{p_{PW}+p_{RW}}\frac{p_{RW}}{p_{PW}+p_{RW}}(p_{PW}+p_{RW}) > \frac{p_{PB}}{p_{PB}+p_{RB}}\frac{p_{RB}}{p_{PB}+p_{RB}}(p_{PB}+p_{RB}).$$

We know that $p_{PW}+p_{RW} > p_{PB}+p_{RB}$ so it is sufficient to show that

$$\frac{p_{PW}}{p_{PW}+p_{RW}}\frac{p_{RW}}{p_{PW}+p_{RW}} > \frac{p_{PB}}{p_{PB}+p_{RB}}\frac{p_{RB}}{p_{PB}+p_{RB}}.$$

Using that $p_{RW}/(p_{PW}+p_{RW}) = 1 - p_{PW}/(p_{PW}+p_{RW})$ (and similarly for blacks) we can rewrite this condition as

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}}\right)^2 - \left(\frac{p_{PW}}{p_{PW}+p_{RW}}\right)^2 > \frac{p_{PB}}{p_{PB}+p_{RB}} - \frac{p_{PW}}{p_{PW}+p_{RW}}.$$

Since whites are richer than blacks, both the left and right hand sides of this expression are positive. The left hand side can be rewritten as

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}} + \frac{p_{PW}}{p_{PW}+p_{RW}}\right)\left(\frac{p_{PB}}{p_{PB}+p_{RB}} - \frac{p_{PW}}{p_{PW}+p_{RW}}\right) > \left(\frac{p_{PB}}{p_{PB}+p_{RB}} - \frac{p_{PW}}{p_{PW}+p_{RW}}\right).$$

Since blacks are on average poorer, the right hand side is positive and we can therefore divide both sides by the expression on the right hand side so that the condition simplifies to

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}} + \frac{p_{PW}}{p_{PW}+p_{RW}}\right) > 1.$$

Since $p_{PB}/(p_{PB}+p_{RB}) > p_{PW}/(p_{PW}+p_{RW})$, we can write $p_{PB}/(p_{PB}+p_{RB}) = p_{PW}/(p_{PW}+p_{RW})+\varepsilon$ for some $\varepsilon > 0$. Hence, we need to show that $p_{PW}/(p_{PW}+p_{RW}) > (1-\varepsilon)/2$. The assumption that poor are in majority implies

$$\frac{p_{PB}}{p_{PB}+p_{RB}}(p_{PB}+p_{RB})+\frac{p_{PW}}{p_{PW}+p_{RW}}(p_{PW}+p_{RW}) > \frac{1}{2}.$$

Substituting $p_{PB}/(p_{PB}+p_{RB})$ we can write this condition as

$$\left(\frac{p_{PW}}{p_{PW}+p_{RW}}+\varepsilon\right)(p_{PB}+p_{RB})+\frac{p_{PW}}{p_{PW}+p_{RW}}(1-p_{PB}-p_{RB}) > \frac{1}{2},$$

which can be further rewritten as

$$\frac{p_{PW}}{p_{PW}+p_{RW}} > \frac{1-2(p_{PB}+p_{RB})\varepsilon}{2}.$$

Since $(p_{PB}+p_{RB}) < 1/2$, we have $(1-2(p_{PB}+p_{RB})\varepsilon)/2 > (1-\varepsilon)/2$, and we have therefore shown what we needed to show.

A.2. Proof of Lemma 3. When whites and blacks are equally rich, we can see from (3.1) and (3.2) that the rich will always prefer zero taxes irrespective of how they identify themselves. Poor identifying with their ethnic identities will prefer the tax rate $1-(y_P+\gamma y)/(1+\gamma)y$ whereas poor identifying with their class will prefer the tax rate $1-y_P/y$. Since the poor are in majority, the median tax rate will be either of these two tax rates. The remainder of the result follows directly from the assumptions that the poor are in majority, blacks are in minority and poor whites are in minority.

A.3. Proof of Lemma 4. We know from Lemma 1 that the assumption that whites are richer than blacks imply that $t^*_{RWW} \leq t^*_{RBB}$ and $t^*_{PWW} \leq t^*_{PBB}$. Combining this with the first part of Lemma 1, we see that optimal tax rates can be ordered in two different ways depending on t^*_{RBB} and t^*_{PWW}:

$$0 = t^*_{RBR} = t^*_{RWR} \leq t^*_{RWW} \leq \left[\begin{array}{c} t^*_{RBB} \leq t^*_{PWW} \\ \text{or } t^*_{PWW} \leq t^*_{RBB} \end{array}\right] \leq t^*_{PBB} \leq t^*_{PWP} = t^*_{PBP} = 1-\frac{y_P}{y}.$$

First suppose that poor whites identify as white. From the inequalities above we see that this implies that rich whites prefer lower tax rates and poor blacks prefer higher taxes than poor whites. Suppose that $t^*_{RBB} \leq t^*_{PWW}$ so that rich blacks always prefer a lower tax rate than poor whites. Since the poor are in majority and poor white are in minority, the median tax rate must be t^*_{PWW}, t^*_{PBB} or $t^*_{PBP} = t^*_{PWP}$, but since $t^*_{RBB} \leq t^*_{PWW} \leq t^*_{PBB} \leq t^*_{PBP} = t^*_{PWP}$, the median must be t^*_{PWW}. If we instead assume that $t^*_{PWW} \leq t^*_{RBB}$, then rich blacks may prefer a higher tax rate than poor whites. However, since we have assumed that blacks are in minority the median tax rate must be t^*_{PWW} also in this case.

Now suppose instead that the poor whites identify as poor. Since the poor are in majority and poor whites are in minority, the median tax rate will be either t^*_{PBB} or $t^*_{PWP} = t^*_{PBP}$. Therefore the median tax rate will be t^*_{PBB} if poor blacks identify as black and $t^*_{PWP} = t^*_{PBP}$ if poor blacks identify as poor.

A.4. Proof of Proposition 3. Let p_B denote the proportion of poor and p_P the proportion of rich. Since whites and blacks are equally rich, $p_{PB} = p_P p_B$, $p_{RB} = (1 - p_P) p_B$, $p_{PW} = p_P (1 - p_B)$ and $p_{RW} = (1 - p_P)(1 - p_B)$. Using these relations the two conditions for ethnic identification (A.1) and (A.2) can be rewritten as

$$PB : \frac{1 - p_B}{1 - p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1 - t)(y_R - y_P)\right),$$

$$PW : \frac{p_B}{1 - p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1 - t)(y_R - y_P)\right).$$

Since whites and blacks both have the average income y and the average income of both ethnic groups is unchanged, the median tax rates for given identities and the status of different identities remain unchanged.

First suppose that poor whites initially identify as white so that the tax rate is $t^*_{PWW} = t^*_{PBB}$ (according to Lemma 3). An increase of p_B implies that the left hand side of the identity choice inequality for poor whites increases which in turn implies that identity choices and hence the tax rate will remain unchanged.

Now suppose instead that both poor blacks and poor whites identify as poor so that the equilibrium tax rate is $t^*_{PWP} = t^*_{PBP}$. An increase in p_B will then decrease the left hand side of the identity choice inequality for poor blacks which implies that they will not change their identity. Since the right hand side is the same for both poor blacks and poor whites and $p_B < 1/2$, poor whites will never identify as white unless the poor black identify as black and so the tax rate remains unchanged.

Finally, suppose that poor whites identify as poor and poor blacks identify as black so that the tax rate is $t^*_{PWW} = t^*_{PBB}$. Increasing p_B might then induce the poor blacks to switch to the poor identity so that the tax rate will be $t^*_{PWP} = t^*_{PBP}$. Alternatively, poor whites may switch to the white identity, but that would leave the tax rate unaffected (recall that Lemma 2 implies that not both black and white can switch identities in this case).

A.5. Proof of Proposition 4. Denote the proportion of blacks by p_B, the proportion of poor among the blacks as $p_{P|B}$ and the proportion of poor among whites as $p_{P|W}$. This implies that $p_{PB} = p_{P|B} p_B$, $p_{RB} = \left(1 - p_{P|B}\right) p_B$, $p_{PW} = p_{P|W}(1 - p_B)$

and $p_{RW} = (1 - p_{P|W})(1 - p_B)$. To see how identity choices are affected, we use these relations to rewrite the inequalities (A.1) and (A.2) as

$$PB: \frac{p_{P|W}}{\left(p_{P|B}\left(\frac{p_B}{1-p_B}\right) + p_{P|W}\right)(1 - p_{P|B})} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right),$$

$$PW: \frac{p_{P|B}}{\left(p_{P|B} + p_{P|W}\left(\frac{1-p_B}{p_B}\right)\right)(1 - p_{P|W})} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right).$$

We see that, for a given tax rate, the poor identity becomes more attractive for poor blacks as the proportion of blacks increases, whereas the white identity becomes more attractive for poor whites.

Since we increase the proportion of blacks but keep the average income of blacks and whites constant, the average income in the population decreases, implying lower optimal tax rates for given identities. This makes the ethnic identities more attractive for both poor blacks and poor whites.

If poor whites initially identify as white, the tax rate is t^*_{PWW}. More blacks will both decrease the tax rate and make the left hand side of the inequality for poor whites higher, which implies that poor whites will continue to identify as white and the tax rate will remain at t^*_{PWW}.

If poor blacks and poor whites identify as poor, the median tax rate is $t^*_{PBP} = t^*_{PWP}$. Since the tax rate decreases, we cannot be certain that poor blacks will continue to identify as poor. If poor blacks switch to the black identity, poor whites may switch to the white identity. In all these different cases, the end effect till be a decrease in the tax rate.

If poor whites initially identify as poor, but poor blacks as black, the median tax rate will be t^*_{PBB}. For given identity choices, the tax rate decreases. Poor black may switch to the poor identity when the proportion of blacks increases, whereas poor whites may switch to the white identity. However, both poor blacks and poor whites cannot change identities since poor blacks identify as black whenever poor whites identify as white. The end effect is therefore not clear in this case.

A.6. Proof of Proposition 5. Denote the proportion of poor by p_P, the proportion of blacks among the poor by $p_{B|P}$ and the proportion of blacks among the rich by $p_{B|R}$. Then $p_{PB} = p_{B|P}p_P$, $p_{RB} = p_{B|R}(1 - p_P)$, $p_{PW} = (1 - p_{B|P})p_P$ and $p_{RW} = (1 - p_{B|R})(1 - p_P)$. Using these relations we can rewrite the conditions for

ethnic identification (A.1) and (A.2) as

$$PB: \left(1-p_{B|P}\right)\left(1+\frac{p_{B|P}p_P}{p_{B|R}\left(1-p_P\right)}\right) > \frac{1}{\beta}\left(\alpha-\frac{\gamma}{\delta}\left(1-t\right)\left(y_R-y_P\right)\right),$$

$$PW: p_{B|P}\left(1+\frac{\left(1-p_{B|P}\right)p_P}{\left(1-p_{B|R}\right)\left(1-p_P\right)}\right) > \frac{1}{\beta}\left(\alpha-\frac{\gamma}{\delta}\left(1-t\right)\left(y_R-y_P\right)\right).$$

By differentiating the left hand sides with respect to $p_{B|P}$ we see that the left hand sides of these inequalities are increasing in $p_{B|P}$ if the following holds:

$$PB: p_{B|P} < \frac{p_P-\left(1-p_P\right)p_{B|R}}{2p_P} = \hat{p}_{B|P},$$

$$PW: p_{B|P} < \frac{1-\left(1-p_P\right)p_{B|R}}{2p_P} = \bar{p}_{B|P}.$$

Note that these thresholds satisfy $\hat{p}_{B|P} < 1/2 < \bar{p}_{B|P}$.

An increase in $p_{B|P}$ implies that average income in the population is unchanged, whereas the income of the black group decreases and the income of the white group increases. This implies that t^*_{PWW} decreases and t^*_{PBB} increases, while $t^*_{PBP} = t^*_{PWP}$ is unaffected.

First suppose that poor whites initially identify as white, and consequently poor blacks identify as black, so that the median tax is t^*_{PWW}. Then the right hand side of the identity choices is decreasing in $p_{B|P}$, so we can be certain that poor whites continue to identify as white if the left hand side is increasing, i.e., if $p_{B|P} < \bar{p}_{B|P}$. If this condition is not satisfied, we cannot say what happens to the identity choice of poor whites and therefore not what happens to the tax rate either.

Now suppose that both poor whites and poor blacks identify as poor. This implies that the tax rate is $t^*_{PBP} = t^*_{PWP}$, which is unaffected by an increase in $p_{B|P}$ so that the right hand sides above are constant. If $p_{B|P} < \hat{p}_{B|P}$, then poor blacks (and possibly also poor whites since $\hat{p}_{B|P} < \bar{p}_{B|P}$) might switch to the ethnic identity. In all cases, this means that the tax rate may decrease if poor blacks (and possibly also poor whites) switch identities.

Finally, suppose that poor whites identify as poor and poor blacks as black so that the tax rate is t^*_{PBB}. This tax rate is increasing in $p_{B|P}$, so the right hand sides of the conditions for ethnic identification increases, which makes the poor identity more attractive. If $p_{B|P} > \bar{p}_{B|P}$, then the left hand sides are decreasing and poor blacks might switch to the poor identity while poor whites continue to identify as poor. In this case, the tax rate will increase, whereas we cannot tell what will happen if $p_{B|P} < \bar{p}_{B|P}$.

A.7. Proof of Proposition 6. As income inequality changes, the only thing that changes in the conditions for ethnic identification (3.3) and (3.4) is the term $(1-t)(y_R - y_P)$ which reflects the relative status of the ethnic identity over the poor identity. Since blacks and whites are equally rich, there are only two tax rates in equilibrium: $t^*_{PWW} = t^*_{PBB} = 1 - (y_P + \gamma y)/(1+\gamma)y$ and $t^*_{PWP} = t^*_{PBP} = 1 - y_P/y$. Clearly, both these tax rates increase with income inequality for given identity choices. Since both the tax rate and income inequality increase, the effect on relative status $(1-t)(y_R - y_P)$ is ambiguous.

Since we keep average income constant, i.e., $\partial y / \partial (y_R - y_P) = 0$, it must hold that

$$\frac{\partial y_P}{\partial (y_R - y_P)} = -\frac{1-p_P}{p_P} \frac{\partial y_R}{\partial (y_R - y_P)}.$$

It is the case that $\partial (y_R - y_P) / \partial (y_R - y_P) = 1$ and this implies that

$$\frac{\partial y_R}{\partial (y_R - y_P)} - \frac{\partial y_P}{\partial (y_R - y_P)} = 1.$$

Combining these two observations we get

$$\frac{\partial y_P}{\partial (y_R - y_P)} = -(1-p_P) \text{ and } \frac{\partial y_R}{\partial (y_R - y_P)} = p_P.$$

Now consider the case when poor blacks identify as black (and perhaps poor whites identify as white). For given identity choices, the tax rate increases. If poor blacks switch to the poor identity, then the tax rate increases further. To see if this can happen, note that the effect on relative status of higher income inequality is given by

$$\begin{aligned}\frac{\partial (1 - t^*_{PBB})(y_R - y_P)}{\partial (y_R - y_P)} &= (1 - t^*_{PBB}) - \frac{\partial t^*_{PBB}}{\partial (y_R - y_P)}(y_R - y_P) \\ &= \frac{y_P + \gamma y}{(1+\gamma)y} - \frac{1-p_P}{(1+\gamma)y}(y_R - y_P).\end{aligned}$$

Rearranging the latter expression shows that relative status is increasing in income inequality if and only if $y_P/y > (1-\gamma)/2$. If this condition is satisfied, then the identity choice of poor blacks remain unchanged. Otherwise, poor blacks might switch to the poor identity. From Lemma 2 we also know that if poor blacks switch to the poor identity, then poor whites will switch to the poor identity as well (unless they already identified as poor)

Now consider the case when both poor blacks and poor whites identify as poor. In this case, the effect on relative status is given by

$$\begin{aligned}\frac{\partial (1 - t^*_{PBP})(y_R - y_P)}{\partial (y_R - y_P)} &= (1 - t^*_{PBP}) - \frac{\partial t^*_{PBP}}{\partial (y_R - y_P)}(y_R - y_P) \\ &= \frac{y_P}{y} - \frac{1}{y}(1-p_P)(y_R - y_P).\end{aligned}$$

This is increasing if $y_P/y > 1/2$. Note that if this condition holds, then relative status at the ethnic tax rate is increasing for all values of γ. If $y_P/y > 1/2$, then poor blacks (and possibly poor whites) might switch to ethnic identities, implying that the net effect on the tax rate is ambiguous. However, if $y_P/y < 1/2$, identity choices remain unchanged and the tax rate will increase.

A.8. Proof of Proposition 7. Let p_P denote the proportion of poor in the population, p_B the proportion of blacks and $p_{P|B}$ the proportion of poor among blacks. This implies that $p_{PB} = p_{P|B}p_B$, $p_{RB} = (1 - p_{P|B})\, p_B$, $p_{PW} = p_P - p_{P|B}p_B$ and $p_{RW} = (1 - p_B) - p_P + p_{P|B}p_B$. Using these relations we can rewrite the conditions for ethnic identification (A.1) and (A.2) as

$$PB : \frac{p_P - p_{P|B}p_B}{(1 - p_{P|B})\, p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)\,(y_R - y_P)\right),$$

$$PW : \frac{p_{P|B}p_B\,(1 - p_B)}{(1 - p_B - p_P + p_{P|B}p_B)\, p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)\,(y_R - y_P)\right).$$

The left hand side for poor blacks is always increasing in $p_{P|B}$ whereas the left hand side for poor whites is increasing if $p_B < 1 - p_P$.

Although the average income in the population is constant when $p_{P|B}$ increases, the income of the black group falls and the income of the white group increases. This implies that t^*_{PBB} increases for given identity choices, t^*_{PWW} decreases and the tax rate associated with the poor identity is unchanged.

First suppose that poor whites identify as white (and thereby poor blacks identify as black). The tax rate is t^*_{PWW} which is decreasing in $p_{P|B}$. As long as $p_B < 1 - p_P$, the poor whites will continue to identify as white. If this is not satisfied, they may change to the poor identity, and the end effect on the tax rate is unclear.

Now suppose that poor whites and poor blacks identify as poor so that the tax rate does not change as a response to an increase in $p_{P|B}$. Then poor blacks may switch to a black identity implying a lower tax rate, and if $p_B < 1 - p_P$ poor whites may switch to the white identity.

Finally, suppose that poor blacks identify as black and poor whites as poor. Then the tax rate is increasing in $p_{P|B}$ for given identity choices. Poor blacks may therefore switch to the poor identity, which would increase the tax rate. In order to be certain that poor whites do not switch identity, we must assume that $p_B > 1 - p_P$.

Appendix B: Extended Black and White Model

In this appendix, we extend the black and white model with income heterogeneity within social classes. Let there be a continuum of agents that are endowed with pre-tax income $y_i > 0$ and let the pre-tax income of blacks and whites be given by the cumulative income distributions $B(y_i)$ and $W(y_i)$, both with support $(0, \max\{y_i\}]$. The distribution of income in the whole population is given by

$$F(y_i) = p_B B(y_i) + (1 - p_B) W(y_i),$$

where $p_B < 0.5$ is the proportion of blacks in the population. Let y denote the average pre-tax income in the population, and let the average income of the four different social categories be denoted by y_B, y_W, y_P and y_R. As in the standard black and white model, the poor are defined as those with pre-tax income below the average income, and we assume that they are in majority, i.e., $F(y) > 0.5$.

All members of a particular type make the same identity choice. The right hand sides of the conditions for ethnic identification are the same as in (3.3) and (3.4) for poor whites and poor blacks, whereas the left hand sides depend on the average incomes of whites and blacks. Note that the assumptions made above imply that we can write the proportion of the four different types as $p_{PB} = p_B B(y)$, $p_{RB} = p_B(1 - B(y))$, $p_{PW} = (1 - p_B) W(y)$ and $p_{RW} = (1 - p_B)(1 - W(y))$. Using these relations, we can rewrite the conditions for ethnic identification of poor blacks and poor whites as

$$\text{(A.3)} \quad PB : (1-t)(y_B - y_P) > \frac{\delta}{\gamma}\left(\alpha(1 - B(y)) - \beta\frac{(1-p_B)W(y)}{p_B B(y) + (1-p_B)W(y)}\right),$$

$$\text{(A.4)} \quad PW : (1-t)(y_W - y_P) > \frac{\delta}{\gamma}\left(\alpha(1 - W(y)) - \beta\frac{p_B B(y)}{p_B B(y) + (1-p_B)W(y)}\right).$$

Since the average incomes in the black and white groups can no longer been expressed as the proportion of different types in the population, these conditions cannot be simplified further. In particular, we cannot use Lemma 2 to rule out the case where poor blacks identify as poor and poor whites identify as white.

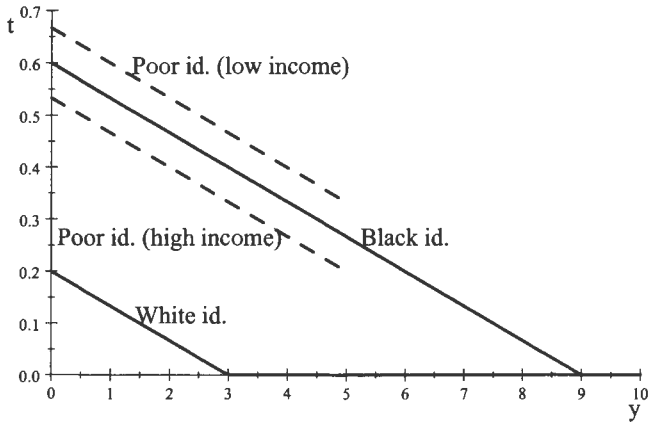
From the optimal tax rate (2.2) we know that the tax rate preferred by an agent with income y_i identifying with group j is given by

$$\text{(A.5)} \quad t^*_{i,j} = \max\left\{1 - \frac{1}{1+\gamma}\frac{y_i}{y} - \frac{\gamma}{1+\gamma}\frac{y_j}{y}, 0\right\}.$$

This implies that rich whites and rich blacks prefer zero tax rates when they identify as rich. In order to see how preferred taxes for the different types are related, note first that the slope with respect to individual income is always the same (unless the preferred tax rate is zero). Figure A1 shows preferred tax rates as a function of individual pre-tax income when $\gamma = 2$, average income in the population is 5 and the average income

of the white and black group is 6 and 3, respectively. The bold lines indicates the preferred tax rate when individuals identify with their ethnic group. The line with higher tax rates corresponds to the tax rates preferred by blacks. In this extended model, the average income of the poor might be higher or lower than the average income of blacks. Figure A1 therefore depict two different cases when the average income of the poor is 2.5 and 3.5. The dashed lines indicates the corresponding tax rates when the poor identify with the poor.

FIGURE A1. Preferred tax rates (interethnic income inequality)



It is clear from Figure A1 that rich blacks may prefer a higher tax rate than poor whites, implying that the identity choice of rich blacks must also be taken into account. Although this could be handled easily in the black and white model (see the proof of Lemma 4), it complicates the analysis of the extended model considerably. There are eight possible combinations of identity choices to take into account. For each of these combinations, there might be up to three median voters, which makes it tedious to solve for the equilibrium in the general case.

Because of the difficulty in studying the equilibria in the general case, we simplify by assuming that blacks and whites have the same average income, i.e., $y_B = y_W = y$, and that the proportion of poor is the same among blacks and whites, i.e., $B(y) = W(y) = F(y)$. Under these assumptions, the conditions for ethnic identification (A.3)

and (A.4) simplify to

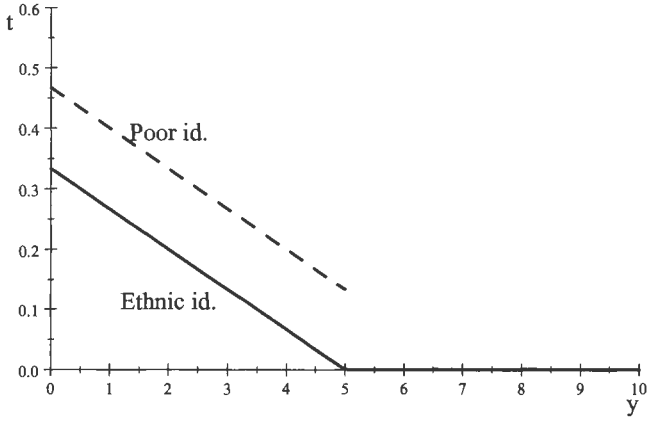
$$PB : (1-t)(y-y_P) > \frac{\delta}{\gamma}(\alpha(1-F(y)) - \beta(1-p_B)), \tag{A.6}$$

$$PW : (1-t)(y-y_P) > \frac{\delta}{\gamma}(\alpha(1-F(y)) - \beta p_B). \tag{A.7}$$

The left hand side is the same for both types, and the right hand side only differs in the cognitive distance to the poor identity. Since whites are in majority, the right hand side is always smaller for poor blacks than poor whites, i.e., whenever poor whites identify as white, poor blacks identify as black. Note that this would not be true if the proportion of poor within the two ethnic groups differed, i.e., if $B(y) \neq W(y)$.

Now consider optimal tax rates under these assumptions. The rich will always prefer zero taxes when identifying as rich, and poor blacks and whites prefer the same tax rate when they identify with their ethnic group for a given level of income. The preferred tax rates may look like in Figure A2.

FIGURE A2. Preferred tax rates (no interethnic income inequality)



As can be seen from Figure A2 and optimal tax rate (A.5), the poor always prefer positive taxes, whereas the rich always prefer zero taxes. Since the poor are in majority, this implies that the median voter (or voters) is poor and that the median tax is determined by the identity choices of the poor. Since poor blacks identify as black whenever poor whites identify as white, there are only three possible combinations of identity choices of the poor in equilibrium. Lemma 5 states these three combinations and the relation between the associated median tax rates.

LEMMA 5. *If whites and blacks have the same average income and proportion of poor, there are three potential equilibrium tax rates satisfying* $t_E^m < t_{PE}^m < t_P^m$*:*

(1) *If poor whites and poor blacks identify as poor, then the median tax rate is* t_P^m*.*
(2) *If poor whites identify as poor and poor blacks as black, then the median tax rate is* t_{PE}^m*.*
(3) *If poor whites identify as white, then the median tax rate is* t_E^m*.*

PROOF. Let the tax rate of poor blacks or poor whites when they identify with the poor be given by $t_P(y_i)$, and let $t_E(y_i)$ denote preferred taxes when they identify with their ethnic group. Using (A.5) we can write these taxes as

$$t_P(y_i) = 1 - \frac{1}{1+\gamma}\frac{y_i}{y} - \frac{\gamma}{1+\gamma}\frac{y_P}{y},$$
$$t_E(y_i) = 1 - \frac{1}{1+\gamma}\frac{y_i}{y} - \frac{\gamma}{1+\gamma}.$$

Note that these functions are continuous and decreasing in personal income. In order to determine the median tax, we need to determine the inverse of these two functions. The income corresponding to a certain optimal tax in the two cases are given by

$$y_i(t_P) = (1+\gamma)(1-t_P)y - \gamma y_P,$$
$$y_i(t_E) = (1+\gamma)(1-t_E)y - \gamma y.$$

This allows us to define the median tax that will prevail for the three possible combinations of identity choices.

(1) *Both poor whites and poor blacks identify as poor.* In this case, both poor blacks and poor whites prefer the tax given by $t_P(y_i)$. Since this function is decreasing in personal income, the median voters (one poor black and one poor white) are therefore given by the tax rate that gets support from exactly half of the population. In other words, the median tax is given by the tax rate t_P^m that solves

$$0.5 = F\left((1+\gamma)(1-t_P^m)y - \gamma y_P\right). \tag{A.8}$$

(2) *Poor whites identify as poor and poor blacks as black.* In this case the preferred tax of poor blacks are given by the bold line in Figure A2, whereas the preferred taxes of poor whites is given by the dashed line. The median voter can either be poor white (if poor whites are in majority), poor black or there can be two median voters – one poor black and one poor white. The median tax is

implicitly given by

$$0.5 = p_B B\left(\max\left\{(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y, 0\right\}\right) + (1-p_B)\,W\left(\min\left\{(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y_P, y\right\}\right). \tag{A.9}$$

Note that if $(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y<0$, then the median voter is poor white, whereas the median voter is poor black if $(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y_P>y$. In the intermediate case, there will be two median voters that prefer the same tax, but generally have different incomes.

(3) *Poor whites identify as white.* In this case poor blacks identify as black, so the taxes preferred by both blacks and whites are given by the decreasing function $t_E(y_i)$. The median tax t^m_E is therefore defined implicitly by

$$0.5 = F\left((1+\gamma)\left(1-t^m_E\right)y-\gamma y\right). \tag{A.10}$$

Since F is increasing in its argument, (A.8) and (A.10) implies that

$$(1+\gamma)\left(1-t^m_E\right)y-\gamma y=(1+\gamma)\left(1-t^m_P\right)y-\gamma y_P. \tag{A.11}$$

Rearranging this we see that

$$t^m_P = t^m_E + \frac{\gamma\left(y-y_P\right)}{(1+\gamma)\,y},$$

i.e., $t^m_P > t^m_E$.

In order to see that t^m_{PE} is between these two tax rates, rewrite (A.8) as

$$0.5 = p_B B\left((1+\gamma)\left(1-t^m_P\right)y-\gamma y_P\right)+(1-p_B)\,W\left((1+\gamma)\left(1-t^m_P\right)y-\gamma y_P\right).$$

Using the relationship (A.11) between t^m_E and t^m_P we can rewrite this as

$$0.5 = p_B B\left((1+\gamma)\left(1-t^m_E\right)y-\gamma y\right)+(1-p_B)\,W\left((1+\gamma)\left(1-t^m_P\right)y-\gamma y_P\right). \tag{A.12}$$

Since the poor are in majority and all agents have positive incomes, it follows that (A.11) is positive, but less than y. We can therefore rewrite (A.12) as

$$0.5 = p_B B\left(\max\left\{(1+\gamma)\left(1-t^m_E\right)y-\gamma y, 0\right\}\right) + (1-p_B)\,W\left(\min\left\{(1+\gamma)\left(1-t^m_P\right)y-\gamma y_P\right\}, y\right). \tag{A.13}$$

Comparing this expression with (A.9) we can show that $t^m_E < t^m_{PE} < t^m_P$. First, suppose by contradiction that $t^m_{PE} \leq t^m_E$. Since F is increasing in its argument, from (A.13) it must be the case that

$$0.5 < p_B B\left(\max\left\{(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y, 0\right\}\right) + (1-p_B)\,W\left(\min\left\{(1+\gamma)\left(1-t^m_{PE}\right)y-\gamma y_P\right\}, y\right),$$

which implies that (A.9) cannot hold and we can conclude that $t^m_{PE} > t^m_E$. Similarly, suppose by contradiction that $t^m_{PE} \geq t^m_P$. Then again using (A.13) and the fact that F is increasing implies that

$$\begin{aligned} 0.5 > p_B B\left(\max\left\{(1+\gamma)\left(1-t^m_{PE}\right)y - \gamma y, 0\right\}\right) \\ + (1-p_B) W\left(\min\left\{(1+\gamma)\left(1-t^m_{PE}\right)y - \gamma y_P\right\}, y\right) \end{aligned}$$

which contradicts (A.9). We conclude that $t^m_P > t^m_{PE} > t^m_E$. □

Now consider an increase in the proportion of blacks in the population. The only effect this will have is to make poor whites more likely to identify as white and poor blacks more likely to identify as poor. The result is therefore very similar to Proposition 3. The only difference is that in the extended model there are three potential median tax rates. This implies that the tax rate may decrease as a result of poor whites switching to the white identity, which could not happen in the simple black and white model without interethnic income inequality (unless the poor whites are in majority).

PROPOSITION 8. *If blacks and whites have the same average income and proportion of poor, then an increase in the black population implies the following for the equilibrium tax rate:*

(1) *If poor whites initially identify as white or both poor whites and poor blacks identify as poor, then the tax rate is unchanged.*
(2) *If poor whites initially identify as poor and poor blacks identify as black, then poor blacks might switch to the poor identity resulting in a higher equilibrium tax rate or poor whites might switch to the white identity resulting in a lower equilibrium tax rate.*

PROOF. From Lemma 5 we know that there are three possible tax rates satisfying $t^m_E < t^m_{PE} < t^m_P$. As p_B increases, the incomes of both groups are constant and it will only affect cognitive distances. From (A.6) and (A.7) it is clear that blacks will feel closer to the poor identity, whereas whites feel more distant to the poor identity. Recall that poor blacks identify as black if poor whites identify as white.

If poor whites initially identify as white (and consequently poor blacks as black), then the prevailing tax rate is t^m_E. An increase in p_B cannot induce poor whites to switch identity, and consequently, poor blacks cannot switch to the poor identity. If both poor blacks and poor whites initially identify as poor, the tax rate is t^m_P and an increase in p_B cannot induce poor blacks to switch identity, which implies that poor whites will not change identity either. Finally, if poor whites initially identify as poor and poor blacks as black, then the initial tax rate is t^m_{PE}. In this case either poor

blacks or poor whites might change identity (but not both), leading to the tax rate t_E^m or t_P^m. □

In the extended model, it becomes considerably more complicated to analyze the case when blacks are on average poorer than whites (as in Proposition 4, 5 and 7). The assumption that blacks and whites have the same average income and the same number of poor simplifies the analysis in three different ways. First, it implies that there are three instead of eight different combinations of identity choices. Second, there are maximally two rather than three median voters for given identity choices. Finally, it implies that we need not consider the two different cases depending on whether poor blacks prefer higher or lower taxes when they identify as black rather than poor.

References

Akerlof, G. A. and Kranton, R. E. (2000), 'Economics and Identity', *Quarterly Journal of Economics* 115(3), 715–753.

Alesina, A. and Angeletos, G.-M. (2005), 'Fairness and Redistribution', *American Economic Review* 95(4), 960–980.

Alesina, A., Baqir, R. and Easterly, W. (1999), 'Public Goods and Ethnic Divisions', *Quarterly Journal of Economics* 115(4), 1167–1199.

Alesina, A., Glaeser, E. and Sacerdote, B. (2001), 'Why Doesn't the United States Have a European-Style Welfare State?', *Brookings Paper on Economics Activity* 2001(2), 187–254.

Alesina, A. and La Ferrara, E. (2005a), 'Ethnic Diversity and Economic Performance', *Journal of Economic Literature* 43(3), 721–761.

Alesina, A. and La Ferrara, E. (2005b), 'Preferences for Redistribution in the Land of Opportunities', *Journal of Public Economics* 89(5-6), 897–931.

Austen-Smith, D. and Wallerstein, M. (2003), 'Redistribution in a Divided Society', Mimeo.

Bénabou, R. and Tirole, J. (2006), 'Belief in a Just World and Redistributive Politics', *Quarterly Journal of Economics* 121(2), 699–746.

Betz, H.-G. (1993), 'The New Politics of Resentment: Radical Right-Wing Populist Parties in Western Europe', *Comparative Politics* 25(4), 413–427.

Bisin, A., Patacchini, E., Verdier, T. and Zenou, Y. (2006), ''Bend It Like Beckham': Identity, Socialization and Assimilation', CEPR Discussion Paper No. 5662.

Corneo, G. and Grüner, H. P. (2000), 'Social Limits to Redistribution', *American Economic Review* 90(6), 1491–1507.

Davis, J. A., Smith, T. W. and Marsden, P. V. (2005), *General Social Surveys 1972-2004*, ICPSR04295-v1, 2005-09-02, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Dincer, O. and Lambert, P. J. (2006), 'Taking care of your own: Ethnic and religious heterogeneity and income inequality', University of Oregon Economics Department Working Paper.

Fernández, R. and Levy, G. (2007), 'Diversity and Redistribution', Mimeo.

Fong, C. (2001), 'Social Preferences, Self-Interest, and the Demand for Redistribution', *Journal of Public Economics* 82(2), 225–246.

Fryer, R. G. and Torelli, P. (2006), 'An Empirical Analysis of 'Acting White", Mimeo, Harvard University.

Gilens, M. (1999), Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy, University of Chicago Press, Chicago.

Glazer, N. and Moynihan, D. P. (1970), *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City*, 2 edn, MIT Press, Boston.

Golder, M. (2003), 'Explaining Variation in the Success of Extreme Right Parties in Western Europe', *Comparative Political Studies* 36(4), 432–466.

Habyarimana, J., Humphreys, M., Posner, D. N. and Weinstein, J. (2007), 'Why Does Ethnic Diversity Undermine Public Goods Provision? An Experimental Approach', *American Political Science Review* (forthcoming).

Kasher, A. and Rubinstein, A. (1997), 'On the Question "Who is a J?": A Social Choice Approach', *Logique et Analyse* 160, 385–395.

Knigge, P. (1998), 'The Ecological Correlates of Right-wing Extremism in Western Europe', *European Journal of Political Research* 34(2), 249–279.

Lind, J. T. (2005), 'Why is there so little redistribution?', *Nordic Journal of Political Economy* 31, 111–125.

Lind, J. T. (2007), 'Fractionalization and the size of goverment', *Journal of Public Economics* 91, 51–76.

Lipset, S. and Marks, G. (2000), *It didn't happen here: Why socialism failed in the United States*, W. W. Norton & Co, New York.

Loury, G. (2002), *The Anatomy of Racial Inequality*, Harvard University Press, Cambridge.

Luttmer, E. F. P. (2001), 'Group Loyalty and the Taste for Redistribution', *Journal of Political Economy* 109(3), 500–528.

Mas-Colell, A., Whinston, M. D. and Green, J. R. (1995), *Microeconomic Theory*, Oxford University Press, New York.

McGann, A. J. and Kitschelt, H. (2005), 'The Radical Right in the Alps', *Party Politics* 11(2), 147–171.

Meltzer, A. H. M. and Richard, S. F. (1981), 'A Rational Theory of the Size of Govern ment', *Journal of Political Economy* 89(5), 914–927.

Miguel, E. and Posner, D. N. (2006), 'Sources of Ethnic Identification in Africa', Mimeo.

Myrdal, G. (1944), *An American Dilemma. The Negro Problem and Modern Democracy*, Harper & Brothers Publishers, New York and London.

Nieuwbeerta, P. and Ultee, W. (1999), 'Class Voting in Western Industrialized Countries, 1945-1990: Systematizing and Testing Explanations', *European Journal of Political Research* 35, 123–160.

Orr, L. L. (1976), 'Income Transfers as a Public Good: An Application to AFDC', *American Economic Review* 66(3), 359–371.

Perotti, R. (1996), 'Growth, Income Distribution and Democracy: What the Data Say', *Journal of Economic Growth* 1(2), 149–187.

Piketty, T. (1995), 'Social Mobility and Redistributive Politics', *Quarterly Journal of Economics* 110(3), 551–584.

Poglia Mileti, F., Tondolo, R., Plomb, F., Schultheis, F., Meyer, M.-H., Hentges, G., Flecker, J. and Mairhuber, I. (2002), 'Modern Sirens and their Populist Songs: A European Literature Review on Changes in Working Life and the Rise of Right-wing Populism', Report from the SIREN project [http://www.siren.at].

Posner, D. N. (2005), *Institutions and Ethnic Politics in Africa*, Cambridge University Press, New York.

Roccas, S. (2003), 'The Effects of Status on Identification withMultiple Groups', *European Journal of Social Psychology* 33, 351–366.

Roemer, J. E. (1998), 'Why the poor do not expropriate the rich: an old argument in new garb', *Journal of Public Economics* 70, 399–424.

Scheve, K. F. and Slaughter, M. J. (2001), 'Labor Market Competition and Individual Preferences over Immigration Policy', *Review of Economics and Statistics* 83(1), 133–145.

Shayo, M. (2005), 'A Theory of Social Identity with an Application to Redistribution', Mimeo, Princeton University.

Soss, J., Schram, S., Vartanian, T. and O'Brien, E. (2001), 'Setting the Terms of Relief: Political Explanations for State Policy Choices in the Devolution Revolution', *American Journal of Political Science* 45(2), 378–395.

Spiegel (2005), 'It's stupid to be afraid', *Spiegel Special International Edition* 2005(7), 21–23.

PAPER 4

Political Polarization and Economic Performance

Erik Lindqvist and Robert Östling

ABSTRACT. We study the effect of political polarization on economic performance using the dispersion of self-reported political preferences as our measure of polarization. Politically polarized countries are poorer and have smaller and inferior governments. The relationship between polarization and economic performance persists when we instrument for polarization. Moreover, polarization of political preferences does not seem to be the result of income inequality per se, divisive economic policies, or higher uncertainty about how to answer survey questions in developing countries. In addition, the levels of political polarization in West and East Germany in 1990 – the year of the reunification – are remarkably similar, indicating that polarization is not endogenous to economic performance. Further, the relationships between polarization of preferences and economic outcomes seem to be stronger in democratic countries, suggesting that part of the effect works through the political system.

1. Introduction

"United we stand, divided we fall"
– Official motto of the State of Kentucky

It is commonly believed that political polarization of the population has detrimental effects on economic performance. For example, the European Council has adopted a policy for "minimizing disparities and avoiding polarization", partly because of its presumed effect on economic development.[1] Economic theory suggests that political polarization may reduce the support for spending on public goods (Alesina et al., 1999), reduce redistribution to the poor (Fernández and Levy, 2007), and increase

[0] A previous version of this paper was entitled "Social Cohesion". We would like to thank Alberto Alesina, Martin Bech Holte, Roberta Dessi, Mikael Elinder, Tore Ellingsen, Justina A. V. Fischer, Christina Fong, Ola Granström, Magnus Johannesson, Ethan Kaplan, Torsten Persson, Daniel N. Posner, Paul Seabright, Jean Tirole, Björn Tyrefors and seminar participants at the ENTER Jamboree 2007 in Mannheim, the Research Institute for Industrial Economics (Stockholm), the 2006 Ratio Colloquium for Young Social Scientists, Stockholm School of Economics and Uppsala University for encouragement and valuable comments. Both authors gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation.

[1] See the first paragraph of the Revised Strategy for Social Cohesion that was approved by the European Council on 31 March 2004.

the scope for rent-seeking if politicians are less constrained by a polarized population (Persson and Tabellini, 2000). Despite the mechanisms suggested by theory and the frequency of appeals to unity by political leaders, there are few empirical studies on the economic effects of political polarization among the population.[2] Moreover, most studies that do consider polarization among the population use indirect measures of political preferences, such as ethnicity or income.

In this paper, we use responses to multiple choice questions about economic policy to derive measures of political polarization at the country level. Our main measure is the standard deviation of responses, but we also consider several alternative measures. All these measures are strong predictors of economic and government performance. Countries that are more cohesive have higher income per capita, more efficient and less corrupt governments, a higher output of public goods and more redistribution of income. The implied relationship between polarization and economic outcomes is strong. For example, consider one of our measures of political polarization based on attitudes toward private ownership of business. Controlling for the mean value of preferences, a one standard deviation increase in this measure predicts a decrease in GDP per capita of 0.51 standard deviations and a decrease in transfers and subsidies as share of GDP of 0.27 standard deviations.

Although we are not able to make a definite test of the direction of causality, there is some evidence suggesting that these relationships reflect a causal effect of political polarization on economic and government performance. First, the relationship between polarization and performance persists when we use ethnic fractionalization and polarization in political interest as instruments for political polarization. Second, the levels of political polarization in West and East Germany in 1990 – the year of the reunification – are remarkably similar, indicating that polarization is stable over time and not endogenous to economic performance. In addition, polarization of political preferences does not seem to be the result of income inequality per se, divisive economic policies, an unwillingness to give extreme answers in certain countries, or higher uncertainty about how to answer survey questions in developing countries. Further, the relationships between polarization of preferences and economic outcomes seem to be stronger in democracies, indicating that part of the effect goes through the political system.

Economic theory suggests a number of mechanisms through which political polarization may affect political and economic outcomes. Alesina et al. (1999) assume that citizens first vote for the amount to spend on a public good and then on the type of public good to provide. As voters in the first stage anticipate the outcome of the

[2] A number of studies have focused on the polarization of political parties. See, for example, Cukierman et al. (1992); Svensson (1998); Frye (2002) and Alt and Dreyer Lassen (2006).

second stage, support for spending on the public good is decreasing in the polarization of preferences over its type. Fernández and Levy (2007) study how political polarization among the poor affects their ability to extract resources from the rich in a model of endogenous party formation. Each taste is represented by an interest group. As taste diversity increases from a low level, redistribution becomes more and more tilted toward special interest groups and general redistribution to the poor goes down. However, at a certain threshold, all special interest coalitions break down and general redistribution to the poor increases. To the extent that political polarization increases electoral uncertainty, a simple probabilistic voting model predicts that this leads to more rent-seeking (see Chapter 4 in Persson and Tabellini, 2000). Glaeser et al. (2005) show how heterogeneous preferences of voters may result in policy divergence of political candidates. If extremism of political candidates have negative economic effects, then this is another mechanism through which polarization of voter preferences can affect economic outcomes. Gradstein and Justman (2002) study the relationship between social cohesion and economic growth. In their model, parents make a trade-off between choosing schooling for their children that foster values which are either close to society's mainstream or close to their own traditional values. Schooling choices of parents in turn affect future productivity since a large social distance between transacting agents is assumed to affect productivity negatively. As a result, social differences diminish and the growth rate increases along the equilibrium path, and social differences are negatively related to growth in the steady state.

At a more general level, a large empirical literature spanning several academic disciplines shows that people who hold similar values are better at coordinating activities and more altruistic toward one another.[3] Political polarization may thus increase transaction costs, lead to less support for redistribution of income and a lower willingness to contribute to public goods.[4]

[3] For example, heterogeneity in preferences leads to lower provision of public goods in experiments (Ledyard, 1995). Relatedly, research within management studies have shown that groups where members hold similar values have fewer intragroup conflicts and better performance (Jehn et al., 1997; Jehn et al., 1999; Jehn and Mannix, 2001), more conversation exchanges (Oetzel, 1998), and a higher degree of social integration (Harrison et al., 1998). Research in social psychology have also shown that attitudinal similarity is one of the strongest predictors of altruism, attraction and friendship, see Newcomb (1961); Byrne (1961, 1971); Suedfeld et al. (1972); Batson et al. (1981); McGrath (1984); Feren et al. (1988) and Chen and Kendrick (2002). As attitudes are partly heritable, the relationship between attitudinal similarity and altruism can be rationalized in terms of evolutionary psychology, see Hamilton (1964); Tesser (1993); Olson et al. (2001) and Park and Schaller (2005). The connection between attitudinal and genetic similarity implies that polarization in political preferences might be related to genetic variation within a society.

[4] This notion is also consistent with the finding that ethnic diversity is associated with lower redistribution and less provision of public goods (see, for example, Alesina et al., 2001; Luttmer, 2001; and Vigdor, 2004).

There are also reasons to expect reverse causality, i.e., that economic outcomes have a causal effect on political polarization. In economic theories of redistribution (e.g. Meltzer and Richard 1981), voters' preferences for redistribution merely reflect their relative position in the distribution of income. As a result, political polarization on issues of redistribution is likely to increase with income inequality. Political polarization may also result from divergence in beliefs about the effects of different policies. For example, Dixit and Weibull (2007) show that if people have heterogeneous priors, agents may interpret signals about the effectiveness of various policies differently and political polarization may result after bad outcomes are realized.[5]

To the best of our knowledge, this paper provides the first investigation of the effect of political polarization on economic performance using a measure of polarization based on self-reported opinions.[6] The paper closest to ours is Gerber and Lewis (2004) who infer individual voter preferences from voting records of binary legislative proposals in California and then calculate a measure of district heterogeneity in voter preferences. They find that legislators are most constrained by the preferences of the median voter in homogeneous districts.[7]

In addition to demonstrating a strong relationship between polarization and economic performance, our paper suggests a caveat to the previous literature on social capital. This literature has argued that trust (Fukuyama, 1995; La Porta et al., 1997; Knack and Keefer, 1997 and 2001) or participation in civic organizations (Putnam, 1993 and 2000) are important determinants of economic and government performance. However, none of these measures are robust once we control for political polarization. In contrast, the estimated effect of political polarization on economic outcomes remains stable when we control for social capital.

Relatedly, we show that the estimated effect of ethnic fractionalization on outcomes is sensitive to political polarization, which indicates that political polarization is one of the channels through which ethnic fractionalization affects government performance.

Finally, this paper suggests that it may be important to control for dispersion or polarization in studies that focus on how the mean value of responses to survey question

[5] Whereas the setup in Dixit and Weibull (2007) implies that people's beliefs will converge in the long run, Acemoglu et al. (2006) develop a related model where political polarization of beliefs might be persistent.

[6] We are not the first, however, to study the dispersion of responses to multiple choice questions at the country level. DiMaggio et al. (1996) use survey data to study changes in dispersion and polarization of attitudes over time. Au (1999) and Au and Cheung (2004) study how variation in job characteristics affects certain social outcomes, such as job satisfaction. There is also a literature within social psychology on the determinants of value consensus (e.g. Shalom and Sagie, 2000).

[7] See also Kalt and Zupan (1990); Goff and Grier (1993) and Bailey and Brady (1998).

relate to economic outcomes. This literature is voluminous and include studies on religion, trust, happiness and beliefs about the fairness of redistribution.[8]

In the next section, we describe the data and how we measure polarization. In Section 3, we test the effect of political polarization on economic performance. Section 4 investigates alternative explanations for the relationship between polarization and economic performance. Section 5 focuses on polarization in other domains than economic policy. Section 6 concludes the paper. Data sources and definitions are provided in the Appendix. Supplementary material is provided in a Web Appendix.[9]

2. Data

We use the standard deviation of responses to multiple-choice questions as our main measure of political polarization, which we calculate using data from the World Values Survey (WVS) 1999-2002 (European Values Study Group and World Values Association, 2004).[10] The WVS is based on face-to-face interviews with between 417 and 6,025 respondents in 81 different countries. For most countries, the WVS contains data from about 1,000 respondents. We excluded Montenegro, Northern Ireland and Serbia from the data, since they are not included as separate entities in the other data sources we use. For each question in the WVS, we calculated the mean and standard deviation of the responses for each country. Thereafter, we excluded all variables with binary responses, fewer than 65 observations or a correlation between the country mean and standard deviation with an absolute value of 0.5 or more. We also excluded constructed indexes, questions about personal characteristics (income, age, etc.) and questions for which the alternatives have no natural ordering. The 36 variables that met these criteria were then further classified into the following five categories depending of their theme of inquiry:[11]

1. Economic policy (5 variables).
2. Confidence in governmental and non-governmental institutions (8 variables).
3. Attitudes toward democracy as a political system (8 variables).
4. Other political questions (7 variables).
5. Personal matters (8 variables).

[8] See, for example, Alesina and Angeletos (2005); Guiso et al. (2006); McCleary and Barro (2006) and Knack and Keefer (1997).

[9] The Web Appendix is available at: http://swopec.hhs.se/hastef/abs/hastef0628.htm

[10] Data for some of the countries in this data set is from the 1994-1998 wave of the WVS. To calculate changes in polarization, we use the cumulative WVS 1981-2004 (European Values Study Group and World Values Association, 2006). See Web Appendix A for details.

[11] The full details about this classification are available in Web Appendix A.

We focus on the questions regarding economic policy. However, we discuss the main results for other question categories in section 5, except for *Personal matters* which we will ignore altogether.[12]

Broadly speaking, the economic policy questions measure various aspects of left and right on a 1 to 10 scale. This means that standard deviations vary between 0 and 4.5, where 0 corresponds to the case where everybody gives the same answer and 4.5 to the case where half of the population says "1" and the other half "10". Table 1 displays the wording of each question.

TABLE 1. Economic policy questions

Left/right (e033)	In political matters, people talk of the left and the right. How would you place your views on this scale, generally speaking? (1 Left to 10 Right)
Equality (e035)	How would you place your views on this scale? 1 means you agree completely with the statement on the left (Incomes should be made more equal); 10 means you agree completely with the statement on the right (We need larger income differences as incentives).
Private (e036)	How would you place your views on this scale? 1 means you agree completely with the statement on the left (Private ownership of business should be increased); 10 means you agree completely with the statement on the right (Government ownership of business and industry should be increased).
Government (e037)	How would you place your views on this scale? 1 means you agree completely with the statement on the left (People should take more responsibility to provide for themselves); 10 means you agree completely with the statement on the right (The government should take more responsibility).
Competition (e039)	How would you place your views on this scale? 1 means you agree completely with the statement on the left (Competition is good. It stimulates people to work hard and develop new ideas.); 10 means you agree completely with the statement on the right (Competition is harmful. It brings out the worst in people).

The questions have been slightly abbreviated, see Web Appendix A for the exact wording.

Though these questions measure different aspects of economic policy, the standard deviations calculated from these questions are rather strongly correlated. As shown in Table 2, the lowest correlation is 0.56 and the highest 0.86.

[12] Most questions in the *Personal matters* category refers to the respondent's own behavior, see Web Appendix A for details.

TABLE 2. Pairwise correlations of standard deviations

	Left/right	Equality	Private	Government	Competition
Left/right	1				
Equality	0.60	1			
Private	0.78	0.75	1		
Government	0.62	0.73	0.86	1	
Competition	0.56	0.77	0.71	0.67	1

Our focus on the standard deviation raises a couple of issues related to the measurement of polarization.

First, the standard deviation is only one of many plausible measures of polarization. Unfortunately, we lack guidance from theory as to which measure is most accurate.[13] To check the robustness of our results, we consider eight other measures of dispersion and polarization. We calculate the polarization measure suggested by Esteban and Ray (1994) for three different parameter values and the kurtosis (see DiMaggio et al. 1996 for further discussion about kurtosis as a measure of bimodality).[14] We also calculate the variance, average absolute deviation, mean difference and the Gini coefficient.[15] Table 3 displays the pair-wise correlations between the standard deviation and the other measures. We use the Government question to calculate these correlations, but the results are very similar for the other economic policy questions, except for Left/right where the correlation between the standard deviation and the polarization measures are lower. The correlations between the standard deviation and the polarization measures from Esteban and Ray (1994) depend entirely on the parameter α. The correlation

[13] Two theories suggest particular measures of polarization. The model by Fernández and Levy (2007) calls for a measure of heterogeneity based on the probability that two randomly matched individuals in the population hold the same opinion. Yet this measure does not resonate well with multiple-choice questions as it treats "4" and "5" on a 1 to 10 scale as two groups as distinct as "1" and "10". The model by Alesina et al. (1999) suggests a measure of polarization based on the median deviation from the median response. However, since there are only a few discrete responses to the questions in the WVS, the median distance to the median is a very crude measure of the variation in our data.

[14] Esteban and Ray's (1994) measure of polarization includes a parameter α which loosely speaking measures the extent of sensitivity to polarization – the higher α is, the more the measure departs from the Gini coefficient. In order to satisfy their axioms, it has to be between zero and approximately 1.6. We calculate their measure when this parameter is 0.5, 1.0 and 1.5.

[15] Average absolute deviation is defined as

$$\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$$

and the mean difference as

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|\,.$$

is lower the larger is this parameter. As it turns out, the main results of the paper are very similar regardless of whether we use the standard deviation or any of the other measures, except for kurtosis which has a weaker relationship with economic performance.

TABLE 3. Correlations with other measures (Government)

Measure	Corr.	Measure	Corr.
Variance	0.996	Average absolute deviation	0.992
Polarization 0.5	0.959	Mean difference	0.980
Polarization 1.0	0.818	Kurtosis	−0.732
Polarization 1.5	0.677	Gini	0.434

Second, all measures of dispersion and polarization we consider treat the ordinal scale of responses to multiple-choice questions as an interval scale, i.e., we assume that the difference between an answer of 1 and 2 is the same as the difference between 5 and 6. Mouw and Sobel (2001) demonstrate that it is possible to measure dispersion without this assumption, but we believe that the cost in terms of difficulties in interpreting the results using their measure is greater than the benefits of using a well-known and simple measure like the standard deviation.

Third, as the standard deviation may be correlated with the mean value of responses, we control for the mean in all regressions. However, the mean value need not perfectly reflect the true mean of preferences if responses are centered on either end of the scale (e.g., 1 or 10 on a 1 to 10 scale). For example, consider two countries with continuously normally distributed underlying distributions of preferences with the same mean. If the mean is above 5.5, the country with the highest standard deviation of the underlying distribution will have a lower observable mean because a larger share of respondents have their answers censored at 10. Hence, the measured standard deviation might be informative about the true mean of preferences even if we control for the measured mean.[16] We use survey questions for which the absolute value of the correlation between mean and standard deviation is below 0.5 to mitigate this problem. The correlations between mean and standard deviation are low for the economic policy

[16] A related concern is that countries with mean values closer to one end point or the other will appear to have a lower standard deviation, since end point censoring reduces variability. We will control for this by including the absolute deviation from 5.5 as an additional control variable. This test is also done to control for "false consensus", see Section 4.4.

questions – the exception being the question regarding private ownership of business with a correlation of 0.43.[17]

Finally, if there is a nonlinear relationship between responses at the individual level and the dependent variable, the variance will have a direct effect on the dependent variable. Since we have no data on individual-level counterparts to the aggregate-level outcomes we consider in this paper, we are not able to test for such nonlinearities and this caveat should be kept in mind when interpreting our results.[18]

What characterizes countries with a high or low degree of political polarization? Table 4 lists the ten countries with the highest and lowest standard deviation in the question about government interventions (Government). Perhaps surprisingly, Pakistan is the country with the lowest level of political polarization. This is not a peculiarity of this particular question – as shown in Table A1 in the Appendix, Pakistan has a very low standard deviation for the other economic policy questions as well. However, Pakistan is also among the countries with the lowest response rates.[19] In the case of the Government question, 37 percent of respondents in Pakistan said they didn't know or gave no answer at all.[20] The other countries on the list are less surprising with three Scandinavian countries among the ten most cohesive and five Latin American countries among the most polarized.[21] A complete list with the standard deviation for each country and question is provided in the Appendix.

We now turn to our measures of economic performance and control variables. The indicators of government performance are the same as La Porta et al. (1999), but we have collected more recent data except for government wages and government employment. The updated data refers to year 2000 in most cases and has been obtained from the same sources. In addition, we have collected data on GDP per capita, the Gini index, country size and population from the World Bank (2005). We obtained data on ethnic polarization and the share of a country's area which is mountainous from Montalvo and Reynal-Querol (2005).[22]

[17] The correlations between mean value and standard deviation for the other polarization measures are: Left/right, 0.15; Equality, -0.14; Government, 0.07 and Competition, 0.22.

[18] The problem of nonlinearities at the individual level and aggregate outcomes is discussed by Deaton (2003) in the context of health and income inequality.

[19] See Table G3 in Web Appendix G.

[20] The report from the person responsible for collecting WVS data in Pakistan does not reveal anything particular except that certain regions of the country couldn't be included in the survey for political and security reasons (for example close to the Afghan border). The data from Pakistan may thus not be fully representative although this is probably not a very serious problem.

[21] It should be kept in mind, however, that we cannot readily compare polarization across continents since we only have data from 78 countries. In addition, although WVS contains many developing countries there is a tendency that larger and more developed countries are more likely to be included.

[22] Montalvo and Reynal-Querol (2005) base their variable of mountainous terrain on work by geographer A. J. Gerard for the World Bank's "Economics of Civil war, Crime, and Violence" project.

TABLE 4. Countries with lowest and highest level of political polarization

Rank	Lowest	Government	Rank	Highest	Government
1	Pakistan	1.92	69	Turkey	3.27
2	Israel	1.99	70	Zimbabwe	3.32
3	Netherlands	2.11	71	Tanzania	3.32
4	Denmark	2.15	72	India	3.43
5	Sweden	2.22	73	Bangladesh	3.43
6	South Korea	2.27	74	Brazil	3.44
7	Norway	2.34	75	Venezuela	3.46
8	Great Britain	2.39	76	Dominican Republic	3.52
9	Taiwan	2.42	77	Mexico	3.55
10	Estonia	2.42	78	El Salvador	3.70

Table 5 shows the partial correlations between the standard deviations and the set of control variables controlling for mean responses. Political polarization is moderately correlated with ethnic fractionalization, but not with religious or linguistic fractionalization. Countries with French legal origin are more polarized, whereas countries with German or Scandinavian legal origin are more cohesive. Socialist and British legal origin are not significantly correlated with polarization. The proportion of Catholics is positively correlated with polarization, whereas countries with many Protestants are more cohesive. In line with the conjecture by Fukuyama (1995), political polarization is strongly negatively correlated with trust.[23] We also check the correlations between polarization and another measure of social capital – the proportion of respondents who participate in at least one civic organization (cf. Putnam, 1993, 2000). This correlation is negative, implying that fewer people participate in polarized countries, but quite weak.[24] There is a strong positive correlation between polarization and income inequality as measured by the Gini index. Polarization is also strongly correlated with distance from the equator – countries close to the equator are more polarized. Some polarization measures are moderately correlated with both land area and size of the population – larger countries (in terms of area and population) are more polarized. We find no significant correlations between population density or mountainous terrain and polarization.

[23] There are only two possible answers to the trust question ("Most people can be trusted" and "Need to be very careful") and a higher value implies lower trust. Hence, we cannot calculate a measure of dispersion based on the trust question. This question is used by Knack and Keefer (1997) and other papers that empirically study trust and social capital.

[24] The modest correlation between participation in civic organizations and political polarization on the aggregate level masks a strong negative correlation between polarization and participation for European countries.

TABLE 5. Partial correlations with control variables

	Left/right	Equality	Private	Government	Competition
Fractionalization					
Ethnic fractionalization+	.34***	.21*	.33***	.29**	.40***
Religious fractionalization	-.10	-.06	-.14	-.13	-.20*
Linguistic fractionalization	.20*	-.01	.12	-.01	.08
Legal origin					
Legal origin: Socialist+	-.08	.05	.07	-.01	-.01
Legal origin: French+	.20*	.15	.13	.30***	.43***
Legal origin: German+	-.26**	-.20	-.27**	-.20*	-.32***
Legal origin: Scandinavian+	-.13	-.18	-.25**	-.31***	-.37***
Legal origin: British+	.10	-.02	.10	-.03	.02
Religious denomination					
Religion: Catholic+	-.04	.33***	.11	.26**	.34***
Religion: Muslim+	.25**	-.14	.18	.10	.16
Religion: Other+	.03	-.08	-.08	-.12	-.14
Religion: Protestant+	-.28**	-.25**	-.33***	-.37***	-.45***
Social capital					
Average trust	.44***	.51***	.58***	.47***	.55***
Participation in civic organizations	-.01	-.12	-.03	-.15	-.24*
Income inequality					
Gini index	.51***	.52***	.46***	.55***	.65***
Geography and demography					
Absolute distance from equator	-.54***	-.33***	-.43***	-.56***	-.53***
Percent mountainous terrain	-.03	-.11	.09	.18	.20
Log of Area (sq. km)	.10	.19	.04	.24**	.28**
Log of total population	.16	.16	.01	.29**	.30**
Log of population density	.02	-.10	-.04	-.03	-.06

Partial correlations with control for mean responses. Variables marked with a plus sign belongs to the set of controls used in the long regression reported in Table 6. One asterisk denotes 10 percent significance level, two asterisks denotes 5 percent significance level and three asterisks denotes 1 percent significance level.

3. Economic and Government Performance

In this section, we analyze whether polarization in preferences for economic policy has predictive power for economic and government performance. We use the same dependent variables as La Porta et al. (1999) and divide them into five categories of

government performance.[25] In addition, we include the logarithm of GDP per capita as a dependent variable. Let y_i denote some economic outcome in country i. For each economic outcome measure, we run the regression

$$y_i = \alpha + \beta Polarization_i + \mathbf{X}_i\gamma + \varepsilon_i, \tag{3.1}$$

where $Polarization_i$ is a measure of polarization in country i and $\mathbf{X}_i$ is a vector of control variables measured at the country level.[26]

There are three potential problems that may bias the estimate of β. First, $Polarization_i$ might be endogenous. In particular, we are concerned that good economic outcomes may reduce the level of polarization, giving rise to a spurious correlation between polarization and bad economic performance. Second, we may get an omitted variable bias if variables which are correlated with both polarization and outcomes are not included in $\mathbf{X}_i$. However, as we are interested in political polarization as a cause of outcomes, we do not want to include variables which are themselves affected by polarization. Third, if polarization is measured with error, our estimate of β will have an attenuation bias toward zero. As these different forms of bias are likely to contradict each other, it is not possible a priori to infer the direction of the total bias.

We run two different specifications for each dependent variable and measure of polarization. In the "short" regression, we only control for the mean response. In the "long" regression, we include controls for legal origin (Socialist, French, German and Scandinavian), share of religious denominations (Muslim, Catholic and Protestant) and ethnic fractionalization.[27] The difference between the short and long regressions is our first test of omitted variable bias. As it is not likely that the control variables in the long regression are affected by political polarization, we should be sceptical to a causal interpretation if the estimated effect of political polarization changes a lot between the short and long regressions. However, adding covariates is likely to reduce the signal-to-noise ratio and thus exacerbate attenuation bias, complicating the interpretation of the difference between the short and long regressions.

Table 6 includes the polarization coefficients from 180 regressions of which 143 are statistically significant at the 10 percent level. In the short regressions, a high level of political polarization is strongly associated with low income per capita. Polarized

[25] We refer to La Porta et al. (1999) for a thorough discussion of the interpretation of these variables.

[26] We have considered using an index combining all economic policy questions instead of reporting the results for the five questions separately. We have abstained from doing so for three reasons. First, there would be fewer observations for this index than for any single question. Second, the results for an index would be less straightforward to interpret. Third, comparing the results for five similar but not identical questions is a sensible first robustness check.

[27] The latter specification is identical to one of the specifications in La Porta et al. (1999) with the exception that we use a different measure of ethnic fractionalization.

countries also have a smaller size and lower quality of government. Most of these relationships are robust to the inclusion of the set of controls in the long regression. The polarization measure that appears to be most robust against the control variables is Government, whereas Left/right appears to be the most sensitive.[28] We also run our measures of economic performance on each of the eight alternative polarization measures with similar results.[29]

We now turn to a more detailed discussion of the results in Table 6.

Economic output. For all economic policy questions, there is a strong negative relationship between polarization and GDP per capita. An increase in polarization by one standard deviation predicts a reduction of GDP per capita by between 0.36 and 0.52 standard deviations, or 54 and 77 percent, in the short regressions and between 0.34 and 0.44 standard deviations, or 51 and 66 percent, in the long regressions. The polarization coefficients are highly statistically significant both in the short and long regressions.

Interference with the private sector. Except for the top marginal tax rate, countries with polarized preferences are more interventionist, with weaker property rights and more regulation that is burdensome for business. The relationship between polarization and property rights is somewhat sensitive to the inclusion of controls, but the relationship with regulation is actually strengthened for two out of five questions. The negative relationship between polarization and the top marginal tax rate is statistically significant in three out of five short regressions, but all coefficients are insignificant in the long regression.

Efficiency. Polarized countries have less efficient governments – there is more corruption, lower bureaucratic efficiency, more tax evasion and higher government wages the higher is polarization. The most robust relationship is between polarization and tax evasion. The results for corruption and bureaucratic delays are somewhat more sensitive to the inclusion of the control variables. The relationship between government wages and polarization is either strengthened or weakened in the long regression depending on the particular question.

[28] Since the Left/right question has the by far the lowest response rates, it is likely that polarization in this regard is measured with more error, giving rise to a larger attenuation bias.

[29] The number if statistically significant coefficients for the other measures are: variance (147), polarization measures with $\alpha = 0.5$ (147), $\alpha = 1.0$ (143) and $\alpha = 1.5$ (136), average absolute deviation (138), mean difference (131), Gini (147) and kurtosis (85). These results are reported in Web Appendix C.

TABLE 6. Economic and government performance

Dependent variable	Left/right		Equality		Private		Government		Competition	
	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
Log of GDP per capita	-1.86***	-1.24***	-1.42***	-1.44***	-1.63***	-1.10***	-1.88***	-1.64***	-2.23***	-1.95***
	(-4.51)	(-3.81)	(-2.75)	(-3.78)	(-3.57)	(-2.95)	(-4.94)	(-4.81)	(-5.57)	(-4.55)
Interference with private sector										
Property rights+	-1.10***	-0.43	-0.96**	-0.67**	-1.30***	-0.63*	-1.42***	-1.07***	-1.56***	-1.11***
	(-3.24)	(-1.50)	(-2.62)	(-2.16)	(-3.72)	(-1.75)	(-4.46)	(-3.36)	(-4.74)	(-3.25)
Regulation	0.40*	0.14	0.58**	0.63**	0.58**	0.45	0.77***	0.77**	0.74**	0.86*
	(1.78)	(0.56)	(2.05)	(2.06)	(2.41)	(1.45)	(2.84)	(2.52)	(2.65)	(1.87)
Top marginal tax rate	-8.07*	-5.71	-3.35	-4.49	-0.97	0.21	-8.37**	-7.49	-9.01**	-3.13
	(-1.98)	(-1.22)	(-0.84)	(-1.02)	(-0.31)	(0.07)	(-2.13)	(-1.63)	(-2.27)	(-0.52)
Efficiency										
Corruption+	1.24***	0.61**	1.07**	0.84**	1.03***	0.67*	1.35***	1.00***	1.69***	1.18***
	(3.78)	(2.12)	(2.55)	(2.44)	(2.73)	(1.79)	(4.18)	(3.27)	(3.97)	(2.81)
Bureaucratic delays+	0.82***	0.38**	0.59**	0.39*	0.39	0.20	0.73***	0.40**	1.15***	0.96***
	(4.15)	(2.14)	(2.42)	(1.87)	(1.69)	(1.08)	(4.57)	(2.18)	(9.15)	(3.80)
Tax evasion+	1.72***	1.34***	1.71***	1.29***	1.37***	1.51***	1.47***	1.22***	1.90***	1.60***
	(6.24)	(3.93)	(5.07)	(3.16)	(5.22)	(3.39)	(6.60)	(3.73)	(6.69)	(3.01)
Government wages	1.53**	1.10	1.33**	1.47**	1.58**	1.59***	1.24**	1.26**	0.99*	0.07
	(2.28)	(1.43)	(2.02)	(2.24)	(2.65)	(2.82)	(2.63)	(2.41)	(1.92)	(0.08)
Output of public goods										
Log of infant mortality	1.52***	1.02***	1.22***	1.23***	1.36***	.99***	1.56***	1.37***	2.04***	1.79***
	(6.00)	(4.55)	(3.27)	(4.75)	(3.89)	(3.35)	(5.55)	(5.90)	(10.95)	(6.98)
Log of schooling	-0.34***	-0.20**	-0.29**	-0.29***	-0.36***	-0.27***	-0.41***	-0.34***	-0.48***	-0.27**
	(-3.58)	(-2.27)	(-2.41)	(-3.31)	(-2.96)	(-2.83)	(-4.06)	(-4.27)	(-5.96)	(-2.06)

Dependent variable	Left/right		Equality		Private		Government		Competition	
	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
Literacy rate	-13.18***	-6.76*	-4.63	-7.78*	-11.44*	-7.94	-11.83*	-10.90**	-13.04***	-8.83**
	(-2.79)	(-1.80)	(-0.73)	(-1.76)	(-1.72)	(-1.66)	(-1.99)	(-2.51)	(-4.58)	(-2.19)
Infrastructure quality	-0.98***	-0.44*	-0.91***	-0.45	-0.46*	-0.23	-0.96***	-0.45**	-1.64***	-1.40***
	(-3.85)	(-1.79)	(-2.96)	(-1.58)	(-1.86)	(-1.00)	(-4.65)	(-2.10)	(-9.07)	(-6.43)
Size of public sector										
Gov. transfers & subsidies	-7.49***	-4.99*	-3.10	-5.40**	-4.38*	-4.27**	-8.73***	-9.65***	-9.33***	-9.45***
	(-3.09)	(-2.00)	(-1.03)	(-2.09)	(-1.70)	(-2.35)	(-3.77)	(-3.65)	(-3.88)	(-3.09)
Gov. consumption	-5.40***	-4.64**	-5.89**	-6.15**	-5.36***	-4.62**	-9.52***	-9.40***	-10.92***	-10.94***
	(-3.25)	(-2.29)	(-2.52)	(-2.46)	(-2.70)	(-2.19)	(-4.53)	(-3.88)	(-5.63)	(-4.12)
Government invest.	-2.19**	-1.20	-0.99	-1.02	-1.85*	-1.34	-0.97	-0.88	-0.81	-1.12
	(-2.50)	(-1.58)	(-0.93)	(-0.98)	(-1.97)	(-1.36)	(-1.13)	(-1.06)	(-0.75)	(-0.89)
Gov. employment	-3.21***	-1.81**	-2.78**	-1.45*	-2.58***	-1.54**	-4.79***	-2.61***	-4.69***	-1.99**
	(-3.99)	(-2.51)	(-2.45)	(-1.98)	(-2.68)	(-2.08)	(-3.69)	(-2.86)	(-3.19)	(-2.08)
Political freedom										
Democracy index	-2.66**	-0.98	-1.37	-1.66	-2.57**	-1.27	-2.43**	-1.97*	-1.93**	-0.85
	(-2.36)	(-0.98)	(-1.07)	(-1.43)	(-2.03)	(-1.09)	(-2.34)	(-1.71)	(-2.55)	(-0.64)
Political rights^{+}	-1.53***	-0.37	-0.76	-0.87	-1.35**	-0.49	-1.54***	-1.20**	-1.43***	-0.86
	(-2.63)	(-0.75)	(-1.16)	(-1.63)	(-2.19)	(-0.93)	(-2.84)	(-2.15)	(-3.15)	(-1.47)

The table reports coefficients and heteroskedasticity robust t statistics for the standard deviation of each question for two different specifications. The left column for each question includes controls for the mean response, whereas the right column also include legal origin, religious denomination and ethnic fractionalization. One asterisk denotes 10 percent significance level, two asterisks denotes 5 percent significance level and three asterisks denotes 1 percent significance level. To facilitate the interpretation, we have switched the sign of the variables marked with a plus sign.

Output of public goods. The output of public goods is significantly lower in polarized countries, in particular regarding infant mortality and the level of schooling. The correlation between infant mortality and polarization is remarkably stable in the long regression. The relationship between polarization and literacy is rather robust, whereas the results for infrastructure quality are quite sensitive to the control variables in the long regression.

Size of public sector. Governments in polarized countries redistribute less income, consume less and employ fewer people. The relationship is robust for redistribution and government consumption, but the effect of polarization on government employment is reduced by about fifty percent in the long regression. The extent of government-owned enterprises and investments (as share of GDP), is slightly larger in polarized countries, but this effect is only statistically significant in two short regressions and in no long regression.

Political freedom. Polarized countries are less democratic and have lower political freedom. However, these relationships are not robust to the inclusion of controls in the long regression.

3.1. Instrument variables. In general, the size of the estimated effects of political polarization on economic outcomes increase when we use two stage least squares to correct for potential endogeneity, omitted variable bias and attenuation bias. We use two different instrument variables for political polarization; ethnic fractionalization and polarization in a WVS question whether "politics is important". The idea behind these instruments is that people who interact a lot tend to develop similar opinions, implying that opinions are more polarized in countries where there is little interaction between groups. We believe that ethnic fractionalization and polarization of political interest may proxy for the extent of such interaction.

In the case of ethnic fractionalization, the first stage regression is

$$Polarization_i = \alpha + \delta Ethnic\ fractionalization_i + \mathbf{X}_i\gamma + \varepsilon_i, \tag{3.2}$$

where $\mathbf{X}_i$ is the same vector of control variables as in the short and long specification of regressions (1), except that ethnic fractionalization is not part of $\mathbf{X}_i$. The second stage regression has the same specifications as regression (1), but ethnic fractionalization is removed from the vector of control variables. The basic first stage regression is the same when we use polarization in political interest as the instrument variable except that ethnic fractionalization is then included in $\mathbf{X}_i$. All results and statistical tests from the instrument variable regressions are reported in Web Appendix D.

Ethnic fractionalization is a quite weak instrument with a first-stage F statistic that is below ten in most cases. Yet the basic pattern from Table 6 persists in the Ethnic IV regressions – political polarization is associated with low economic and government performance and a small public sector. Most of the estimated coefficients are several times larger than the OLS estimates, in particular regarding GDP per capita, infant mortality and transfers and subsidies. One explanation for this result is that the instrument solves attenuation bias due to measurement error in political polarization. However, since standard errors are larger, fewer coefficients (83) are statistically significant at the ten percent level. Except for the Competition question, a Durbin-Wu-Hausman test rejects the null hypothesis of equal probability limits in the Ethnic IV and OLS estimations for GDP per capita, property rights, infant mortality and political rights.[30] As the Ethnic IV estimates are larger, rejection of the null hypothesis suggests that OLS underestimates the true causal effect of political polarization on economic performance. In general, the null hypothesis is not rejected for the other outcome measures.[31]

For ethnic fractionalization to be a valid instrument, it should not affect economic outcomes through other channels than the polarization of political preferences. This assumption is far from innocuous. For example, ethnic groups might constitute a basis for forming coalitions that engage in rent seeking or violent conflict. However, as argued by Horowitz (1985), the relationship between ethnic fractionalization and incidence of civil wars and conflicts may not be monotonic. In particular, Montalvo and Reynal-Querol (2005) argue that it is ethnic *polarization* rather than ethnic *fractionalization* that affects the incidence of civil wars. We therefore include a measure of ethnic polarization in the IV regression as a robustness check.[32] The estimated effects are similar compared to the case without controls for ethnic polarization, but ethnic fractionalization is a weaker instrument in the first stage and the second stage estimates are somewhat less precise with 76 statistically significant coefficients. However, this effect is largely attributed to the smaller number of observations for ethnic polarization, and not an effect of controlling for polarization per se.

As the second instrument, we use the standard deviation in the question whether "politics is important". At the individual level, *Politics important* is highly correlated with "interest in politics" (correlation coefficient 0.57) and how often one discusses political matters (correlation coefficient 0.41). However, the individual level correlations between *Politics important* and the economic policy questions are relatively weak,

[30] This test is based on IV estimates without heteroskedasticity-robust standard errors.

[31] The null hypothesis is rejected six times for government transfers and subsidies, but not more than four times for any other measure of economic performance.

[32] We use the ethnic polarization measure of Montalvo and Reynal-Querol (2005).

ranging from −0.052 to +0.040. We take this as an indication that *Politics important* captures individual engagement in politics rather than a particular political position. Whereas the economic policy questions capture polarization in political preferences, the instrument should measure polarization in political engagement.

The standard deviation of *Politics important* is a strong instrument, the F statistic is above ten in all first stage regressions and often substantially so. The coefficients are more precisely estimated compared to the ethnic fractionalization instrument and 114 coefficients are statistically significant. The size of the polarization coefficients in the second stage are generally in between the OLS estimates and the IV estimates with ethnic fractionalization.[33] For example, whereas a one standard deviation increase in the country level standard deviation of Private predicts an increase in infant mortality by .63 standard deviations in the short OLS and .46 standard deviations in the long OLS regressions, the corresponding figures for the IV regressions are .87 and .66 for *Politics important* and 1.48 and 1.39 for ethnic fractionalization. A Durbin-Wu-Hausman test rejects the null hypothesis for infant mortality in all ten regressions, suggesting that OLS underestimates the true effect of polarization in this case. The null hypothesis is not rejected for more than six regressions for any of the other performance measures.

Including extra controls for the mean value in *Politics important* or participation in voluntary organizations to control for the level of civic engagement does not change the results significantly.[34] Still, we cannot rule out that polarization in political interest has an independent effect on economic performance, and the results from these IV regressions should be interpreted with caution.

Using both ethnic fractionalization and polarization of political interest as instruments gives coefficient estimates which are close to what we obtained for polarization in *Politics important*. As the estimates are more precise when we use both instruments, there are 129 statistically significant polarization coefficients in this case. Infant mortality is the only performance measure for which we persistently reject the null hypothesis in a Durbin-Wu-Hausman test. An over identification-test (Hansen's J statistic) rejects the null hypothesis that the two instruments have equal probability limits in eight out of ten regressions for GDP per capita and property rights. The null hypothesis is not rejected more than four times for any other performance measure, and for more than half of the performance measures not more than once.

[33] Adding the other questions on political interest and participation in the WVS as instruments give stronger results of polarization on quality of government. However, these additional instruments have limited explanatory power in the first stage and are often not statistically significant once we include *Politics important*. To avoid a weak instruments problem (Bound, Jaeger and Baker, 1995), we therefore chose only to stick with *Politics important*.

[34] There were 112 statistically significant coefficients when we controlled for the mean value and 116 statistically significant coefficients when we controlled for participation in civic organizations.

3.2. Stability of polarization. An alternative way of answering the question whether political polarization is endogenous to quality of government is to study "natural experiments" in policy, i.e., shifts in policy which are not affected by polarization. We consider two such cases. First, we compare the degree of political polarization in immigrant groups in the U.S. with polarization in their country of origin. We use five questions about economic policy in the General Social Survey to calculate the level of political polarization within different immigrant groups. These questions are not the same as in the WVS, but they all concern economic policy on a left to right scale. It turns out that polarization among immigrant groups is almost unrelated to polarization in the respective countries of origin. This suggests that the degree of political polarization may not be stable enough for it to survive among immigrants in a new country. However, immigrants are self-selected and not a random sample of the population in their country of origin. In addition, we have very limited number of observations for each group of U.S. immigrants, implying that the polarization measures are very noisy. There are also few immigrant groups that can be matched with specific countries, which leaves us with a small number of observations to estimate the relationship. The details of the analysis of immigrant groups in the U.S. can be found in Web Appendix E.

We also compare the degree of political polarization in East and West Germany in 1990, the year of the reunification. These measures are not biased by self-selection and, as each standard deviation is based on more than 1,200 observations, they are more precise.[35] Moreover, there is obviously much less influence of people from other countries in the German case. Table 7 shows the degree of political polarization for Western and Eastern Germany in 1990 together with the countries with the lowest and highest level of political polarization.The degree of political polarization in Western and Eastern Germany is remarkably similar. The German case seems to support the notion that political polarization is not affected by economic outcomes. However, if political polarization is endogenous with respect to both income inequality and growth, then these two effects may cancel each other out since Western Germany was richer, but Eastern Germany had lower income inequality (Biewen, 2000).

As another test of the stability of political polarization, we consider the intertemporal correlations using data from previous waves of the WVS (1981-1984, 1989-1993 and 1994-1998). As shown in Table 8, the correlations between the old polarization measures and those from WVS 1999-2002 vary between 0.51 and 0.83. These results suggest that the degree of political polarization is relatively stable over time, although we only have data for one polarization measure for a longer time span than a decade.

[35] Of course, Germany only constitute one observation in a cross-country setting.

TABLE 7. Political polarization in Germany in 1990

Left/right			Equality			Private		
Rank	Country		Rank	Country		Rank	Country	
1	Slovenia	1.63	1	Netherlands	2.12	1	Netherlands	1.79
6	West Germany	1.69	14	East Germany	2.60	4	West Germany	2.04
8	East Germany	1.77	17	West Germany	2.75	12	East Germany	2.26
39	Brazil	2.79	42	South Africa	3.61	42	Brazil	3.28

Government			Competition		
Rank	Country		Rank	Country	
1	Sweden	2.20	1	Czech republic	1.77
11	West Germany	2.65	9	West Germany	2.05
31	East Germany	2.93	14	East Germany	2.22
42	Malta	3.48	41	Turkey	3.09

Data for former East and West Germany is from 1990, and from 1989 to 1993 for the other countries.

TABLE 8. Intertemporal correlations with 1999-2002

	1981-1984		1989-1993		1994-1998	
	Correlation coefficient	Obs.	Correlation coefficient	Obs.	Correlation coefficient	Obs.
Left/right	.56	14	.62	35	.80	37
Equality			.51	31	.73	35
Private			.83	29	.75	33
Government			.56	39	.74	39
Competition			.64	38	.72	37

We exclude Argentina and Malta from the first wave as they are extreme outliers for this intertemporal correlation. See Web Appendix A for further details of how the intertemporal correlations were calculated.

To test if changes in measured polarization partly reflect measurement error, we regress change in polarization during the 1990's on polarization in 1990. If polarization is measured with error, we would expect a negative relationship between these two variables. As shown in Table A3 of Web Appendix A, the coefficient has a negative sign in all regressions and is statistically significant in four out of five regressions. This result is stable to controlling for economic growth during the 1990's. Hence, it is likely that the correlation coefficients in Table 8 understate the durability of political polarization and that the estimates in Table 6 are affected by attenuation bias.

3.3. Heterogeneous effects. A further test of causality is to consider heterogeneous treatment effects with respect to democratic development. Arguably, the politico-economic arguments for why polarization affects outcomes are more relevant if voters can influence policies. In contrast, the theoretical arguments for why economic outcomes may affect polarization do not hinge on the level of democratic development.

We classify the 33 countries with a democracy score of 9 to 10 as "strong" democracies and the 39 countries with a score of 0 to 8 as "weak" democracies. We chose this particular cutoff in order to get roughly half of the countries in each group. As illustrated by Figure 1 to 4, the relationship between political polarization and economic outcomes seems to be stronger for the strong democracies.

FIGURE 1. Infant mortality, strong dem.

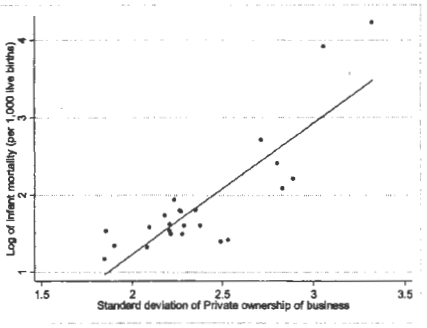


FIGURE 2. Infant mortality, weak dem.

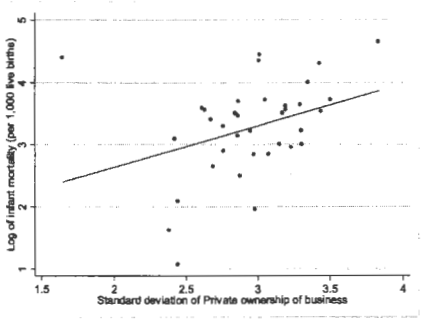


FIGURE 3. Redistribution, strong dem.

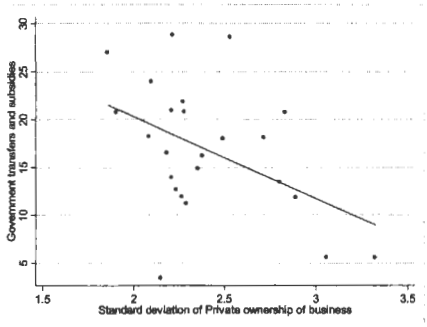
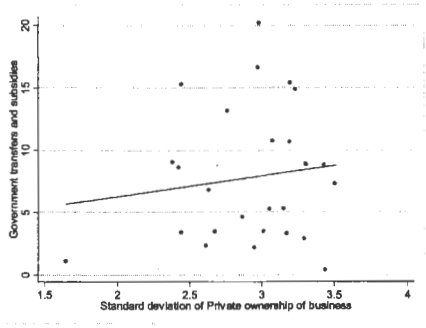


FIGURE 4. Redistribution. weak dem.



More formally, we include an interaction term between political polarization and strong democracy together with a strong democracy dummy in regression (1). The results reported in Web Appendix F indicate that democratic countries have a stronger relationship between political polarization and most of the measures of the size and

quality of government, in particular for GDP per capita, interference with the private sector, efficiency and redistribution. However, neither the common effect, nor the difference between democratic and non-democratic societies, are perfectly consistent across questions. Moreover, even when the results are consistent, they are often not statistically significant.[36]

These results indicate that political polarization affects outcomes through the political process. In particular, the consistent results for property rights, regulation, corruption and bureaucratic delays fit rather well with the finding in Bailey and Brady (1998) and Gerber and Lewis (2004) that legislators are less constrained by voters' preferences in heterogeneous districts. An important caveat to this conclusion is that measurement error in political polarization may vary systematically depending on the level of democratization. For example, it might be easier to collect a representative sample of the population in democracies.

4. Alternative Explanations

In this section, we consider a number of alternative explanations for the relationship between polarization of political preferences and economic performance.

4.1. Economic growth. The negative effect of political polarization on the size and quality of government generally becomes statistically insignificant if GDP per capita is included in the long regressions, suggesting that countries with polarized preferences do not have worse government performance controlling for how poor they are. As GDP per capita is likely to be a function of both quality of government and political polarization, it is hard to draw any strong conclusions regarding the total effect of polarization on government performance from this result.

To better understand the relationship between polarization and growth, we consider changes in political polarization and economic growth over time. First, we run a regression of change in polarization during the 1990's on the logarithm of GDP 1990 and growth during the 1990's. As shown in Table 9, economic growth does not predict changes in political polarization in this regression, implying that there is little co-movement between polarization and growth. The level of GDP per capita is statistically significant in two out of five regressions, indicating that countries which were prosperous in 1990 experienced a decrease in political polarization during the 1990's.

[36] That fewer coefficients are statistically significant is partly due to the loss of statistical power from decomposing one effect into two different, additive effects (i.e., the common effect of polarization and the specific effect of polarization in democracies).

Second, we run a regression of change in polarization during the 1990's on the logarithm of GDP per capita in 1980 and economic growth 1980-1990. In this case, GDP per capita in 1980 has a negative sign and is statistically significant in three regressions (Equality, Private and Competition). Economic growth is statistically significant with a negative sign in two regressions (Private and Competition), indicating reverse causality. However, these results are based on very small samples and not perfectly consistent across questions.

An important question for this paper is whether the effect of growth on polarization, if it exists, is persistent or temporary. If the effect is temporary, it cannot explain the strong relationship between polarization and outcomes. Unfortunately, to test this would require a longer data series than is currently available.[37]

TABLE 9. Dependent variable: Change in political polarization, 1990's

Model	Variable	Left/right	Equality	Private	Government	Competition
1	Log GDP per	-0.04	-0.13***	-0.06	-0.04	-0.09***
	capita 1990	(-1.28)	(-3.40)	(-1.56)	(-0.83)	(-2.77)
	Growth 1990-2000	-0.22	0.04	0.03	-0.01	-0.07
		(-1.26)	(0.22)	(0.30)	(-0.09)	(-0.55)
	Obs	35	31	29	39	38
2	Log GDP per	-0.05	-0.11**	-0.11**	-0.05	-0.10*
	capita 1980	(-1.41)	(-2.66)	(-2.61)	(-0.73)	(-2.39)
	Growth 1980-1990	-0.29	0.01	-0.43*	-0.15	-0.41*
		(-1.27)	(0.02)	(-2.04)	(-0.28)	(-1.73)
	Obs	29	25	24	32	31

Robust t statistics in parentheses. One asterisk denotes 10 percent significance level, two asterisks 5 percent significance level and three asterisks 1 percent significance level.

4.2. Income inequality. As political polarization is strongly related to income inequality at the aggregate level, one might ask if polarization of political preferences is just a reflection of the divergence in people's economic self-interest. We check for this possibility by creating an "income-adjusted" measure of political polarization.[38] For each country j we run the individual question answers on the respondents' income.[39] For example, the regression in the case of the Competition question is

$$Competition_{ij} = \alpha_j + \beta_{1j} Income_{ij} + \varepsilon_{ij},$$

[37] The model by Dixit and Weibull (2007) suggests a temporary relationship between bad economic outcomes and polarization.

[38] All results regarding the income-adjusted measure of polarization are reported in Web Appendix B.

[39] Note that the 1 to 10 scale of income is specific to each country.

where all variables refer to respondent i. We then calculate an "income-adjusted opinion" for each respondent by subtracting the difference between his predicted opinion and the mean predicted opinion from his actual answer. In the case of Competition, we calculate the income-adjusted opinion as

$$Adj.Competition_{ij} = Competition_{ij} - \widehat{\beta}_{1j}\left(Income_{ij} - \overline{Income_j}\right),$$

where $\overline{Income_j}$ is the average income in country j. Finally, we calculate a new polarization measure for each country using *Adj.Competition*. The income-adjusted polarization measures are very similar to the ordinary polarization measures. All correlation coefficients between the standard and income-adjusted polarization measures are above 0.99. Not surprisingly, re-running the regressions of government performance using the income-adjusted standard deviations yields very similar results.[40] The reason for the small differences between the income-adjusted and ordinary polarization measures is found in the low explanatory power of stated income for individual preferences.[41] A potential explanation for this result is that income is measured with error, giving rise to attenuation bias. However, if permanent income is truly an important determinant of political opinions, we would expect individuals' responses to the different economic policy questions to be strongly correlated within countries. In fact, most of the correlations are weak, indicating that the economic policy questions are not easily divided on a single left to right political scale, at least not one that holds for all countries.[42]

The results for polarization are also relatively robust to including the Gini coefficient as a control variable in the basic regression (96 statistically significant coefficients).[43] Since the Gini coefficient is highly correlated with political polarization, and polarization is likely to be measured with error, including the Gini index probably increases in attenuation bias. Another problem is that the Gini index is based on consumption data for developing countries, whereas income net of employer taxes is used for developed countries (Luxembourg Income Study). This implies that the Gini index is affected by taxation and redistribution, thereby spuriously reducing the estimated effect of political polarization. An indication that this is a real problem is that the relationship that is most sensitive to the inclusion of the Gini index is the correlation between political polarization and the size of the public sector. In contrast, the results for GDP

[40] As educational attainment might be a better proxy for permanent income than current income, we also calculate adjusted standard deviations where we include educational attainment (on a 1 to 3 scale) as a regressor in the first stage above. This adjusted measure of heterogeneity is also highly correlated with our standard measure with correlation coefficients of 0.99 for the economic policy questions.

[41] See Web Appendix A for a discussion of individual-level determinants of opinions on the economic policy questions in the WVS.

[42] See Web Appendix A.

[43] See Web Appendix B.

per capita, interference with the private sector, tax evasion and infant mortality are generally not sensitive to the Gini coefficient.

4.3. Divisive economic policies. Since the economic policy questions often relate to the current situation in a particular country, political polarization may be a direct effect of the divisiveness of policy. For example, suppose that the government in a certain country redistributes income from group X to group Y. If people in this country think of "redistribution" as redistribution from X to Y, then measured preferences for "redistribution" may be polarized, even if preferences are homogenous regarding some other redistribution scheme (say from the rich to the poor). If countries in which the government pursues divisive policies also have low quality of government, there will be a spurious correlation between political polarization and economic performance. However, as shown above, the results for economic and government performance are similar for all five economic policy questions. For political polarization to be a direct consequence of divisive economic policies, the divisive policies must thus shape opinions on all these questions. If divisive policies cause polarization, we would therefore expect the responses to the economic policy questions to be strongly correlated at the individual level in countries where polarization is high. In reality, we have the opposite case. As shown in Web Appendix G, the correlation between individual-level responses is much higher in countries with a low level of political polarization.[44]

4.4. Survey data. A potential problem in using survey data is that people in certain countries are uncertain about their political preferences. For example, people in developing countries may be more likely to randomize their responses, giving rise to a spurious correlation between political polarization and poverty.[45] This could also be the case if uncertain respondents choose particular focal values.[46] We perform three different tests to check for this possibility. The results from all tests in this section are reported in Web Appendix G.

First, we include the country response rate as a control variable in regression (1). The idea is that in countries where many people are uncertain about their preferences, more people will also state that they "don't know" what they think or not answer the

[44] In addition, the view that political polarization is a consequence of divisive policies does not fit well with the result that the effect of political polarization is stronger in democracies.

[45] This correlation is not necessarily positive as uniform randomization on a 1 to 10 scale implies a standard deviation of approximately 2.63, far from the theoretical maximum of 4.5 (when half of the respondents answer 1 and the other half 10).

[46] If uncertain respondents have the same focal value, we underestimate the true polarization of preferences. If uncertain respondents instead randomize between extreme values, we overestimate it.

question at all. The results remain very similar to the base case with 149 statistically significant polarization coefficients.

Second, we calculate the country-level correlations between individuals' responses to the economic policy questions. For each question, we then calculate the country average of the absolute value of the correlations with the other questions. This measure captures the extent to which question responses follow a certain pattern, and we expect it to be lower the more uncertain people are about their political preferences. Political polarization remains a robust predictor of economic performance with 130 statistically significant coefficients when we include this measure in regression (1).

Third, as Pakistan is the country with the lowest response rates, we run the regressions with Pakistan excluded from the data set. This gives us somewhat stronger results with 159 statistically significant coefficients. In particular, removing Pakistan strengthens the correlations between polarization and literacy, democracy and political rights.

A related problem with survey data is that people might not want to deviate too much from the opinions of others. A respondent who wants to minimize the maximum deviation from other respondents will answer five or six provided he does not know the others' responses. If so, countries where people are concerned with consensus would have their responses centred around 5.5. Hence, we include the absolute deviation from 5.5 as a control variable in regression (1) as a rough way of controlling for "false consensus". This is also a rough way of controlling for a correlation between the standard deviation and the true mean of preferences due to censoring of the data, as discussed in Section 2. The results in this robustness check are almost exactly the same as in the basic specifications. Another indication that censoring is unlikely to be quantitatively important is that there is no systematic relationship between the mean values (in terms of left or right) and economic performance in the basic specification of regression (1).

4.5. Geography and demography. We also run regression (1) separately including percent mountainous terrain and the logarithm of country area, total population and population density.[47] The results are similar to the basic regressions with at least 130 statistically significant coefficients. In contrast, controlling for absolute distance from the equator cuts the number of statistically significant coefficients to 74. However, there is considerable heterogeneity among the different measures of economic performance in this regard. Whereas the results for output of public goods (except for infant mortality), size of the public sector and political freedom are sensitive to the inclusion

[47] See Web Appendix H for all results discussed in this section.

of absolute distance from the equator, the results for GDP per capita, interference with the private sector and government efficiency are more stable. The results for property rights and tax evasion are actually strengthened slightly. Moreover, we are concerned that distance from the equator is not a strictly exogenous variable since the decision to include it is based on the knowledge of a strong correlation with economic performance.

4.6. Social capital and ethnic fractionalization. A large literature has argued that the level of trust in a society is an important determinant of economic performance.[48] As polarization is strongly negatively correlated with trust, a concern is that polarization merely proxies for trust. However, polarization remains a strong predictor of economic performance with 114 statistically significant coefficients when we include the average level of trust in regression (1).[49] In contrast, the estimated effect of trust on outcomes is sensitive to polarization. Not controlling for polarization, 13 trust coefficients are statistically significant in the specification without any control variables, and 5 when we include the same set of control variables as in the long specification of regression (1). Countries with a high level of trust are richer and have larger and better performing governments. Including our polarization measures and mean values reduces the number of statistically significant trust coefficients substantially. Depending on which polarization measure we control for, between 1 and 6 trust coefficients are statistically significant in the short regression and between 1 and 4 in the long regressions. Moreover, the signs of the trust coefficients are often both negative and positive for the same outcome measure.

The estimated effect of political polarization is also robust to controlling for civic engagement, which along with trust is often considered an important part of social capital (cf. Putnam, 1993, 2000). Including the percentage of people who participate in at least one civic organization in regression (1) strengthens the results for political polarization, though this effect is entirely due to sample selection. In general, civic engagement has a weak positive effect on economic performance in a regression without controls, and a weak negative effect when we include the same set of control variables as in the long regression (except polarization and mean value). Controlling for political polarization reduces the positive effect of civic participation in the short regressions and strengthens the negative effect in the long regressions.

To check the sensitivity of political polarization with respect to ethnic fractionalization, we re-run all long regressions excluding ethnic fractionalization as a regressor.

[48] See, for example, Fukuyama, (1995); La Porta et al., (1997) and Knack and Keefer, (1997, 2001).
[49] See Web Appendix I for all results discussed in this section.

The size of the coefficients only change slightly, though there are 72 statistically significant coefficients in the long regression without ethnic fractionalization compared to 63 if ethnic fractionalization is included. None of the coefficients change sign as we drop ethnic fractionalization.

If political polarization is the main channel by which ethnic fractionalization affects public goods provision (as suggested by for example Alesina et al., 1999), we would expect the coefficient on ethnic fractionalization to be sensitive to the inclusion of political polarization. In the long regression without the mean and standard deviation, ethnic fractionalization has a stable negative effect on performance, with twelve out of eighteen coefficients being statistically significant at the ten percent level. Including the mean and standard deviation reduces the effect of ethnic fractionalization on government performance substantially. Controlling for political polarization, there are only three economic outcomes for which ethnic fractionalization is significant at the five percent level in at least three of the five long regressions (GDP per capita, less secure property rights and higher infant mortality).

5. Other Questions in the WVS

We also run regression (1) using the other question categories as measures of political polarization.[50]

The results for the questions about confidence in governmental and non-governmental institutions are similar to those for economic policy, though, in general, polarization in economic policy is more robust to the set of controls in the long regression. However, polarization in confidence is a more robust predictor of democratic development.

Polarization in preferences for democracy as a political system is relatively weakly related to quality of government. Yet there is one interesting exception to this rule – the question whether having the army rule is a good or a bad thing has the same properties as for the economic policy questions.

The category of "other political questions" is, in itself, more heterogeneous than the others. It includes questions concerning, among other things, the justifiability of divorce, willingness to pay for preventing pollution of the environment and immigrant policy. These questions also differ in their relation to the measures of quality of government. Some questions show the same pattern as the economic policy questions. For example, polarization in the ranking of "the political system as it was before" is a strong predictor of economic performance. In contrast, polarization in "willingness

[50] See Web Appendix J for the results for the other question categories.

to pay for protecting the environment" is almost unrelated to measures of quality of government.

6. Conclusions

This paper has shown that polarized countries have lower economic and government performance and smaller public sectors. This relationship is robust to a large set of control variables and the use of other measures of polarization than the standard deviation. Moreover, we found no evidence that the correlation between political polarization and economic performance is a direct consequence of income inequality, divisive economic policies, false consensus, or an inability to understand the survey questions in developing countries. However, we cannot rule out the possibility that cohesive preferences are caused by economic growth. In fact, we believe that both directions of causality between cohesive preferences and good outcomes are plausible. Good economic outcomes may generate convergence in opinions which in turn generate good outcomes. Bad outcomes might generate polarized preferences which in turn have a detrimental effect on outcomes.

A priority for future research is to better identify the underlying causal mechanism. In order to do this, it would probably be useful to use panel data on political polarization, but it would also be interesting to examine if political polarization can explain variation in performance within countries. A definite test of the causal mechanism, however, requires some kind of exogenous variation in political polarization that does not have an independent effect on economic outcomes. Unfortunately, it is very difficult to come up with an instrument that only affects economic performance through political polarization.[51]

The relationship between political polarization and economic and government performance is relevant for policy. If political polarization has a causal effect on economic performance, then policies that foster homogeneity might improve welfare. One such policy instrument is the educational system to the extent that it plays a role in shaping the values of future citizens (as suggested by Gradstein and Justman, 2002).

[51] One potential instrument could be a randomized propaganda campaign, although it is probably difficult to find such campaigns that are both exogenous and successful enough to change people's political preferences.

Appendix A

TABLE A1. Political polarization in economic policy by country

Country	Left/right	Equality	Private	Government	Competition
Albania	2.75024	2.47269	2.42767	2.77046	2.02929
Algeria	2.58663	2.57152	3.19376	3.01380	
Argentina	2.13119	3.25348	3.07496	3.15056	3.18853
Armenia	2.25541	2.77895	2.84538	2.49102	2.53844
Australia	1.79677	2.55512	2.29370	2.61804	2.11603
Austria	1.64702	2.56862	2.10130	2.56659	2.01197
Azerbaijan	2.38699	2.92124	3.01014	2.79811	2.55140
Bangladesh	2.55723	2.85097	3.35134	3.43077	2.47480
Belarus	1.98132	2.94756	2.69110	2.95749	2.48820
Belgium	1.86787	2.92396		2.72886	2.60430
Bosnia and Herzegovina	1.99338	2.60381	2.65671	2.86548	2.14744
Brazil	2.89429	3.44113	3.19661	3.44391	2.92267
Bulgaria	2.50051	3.00508		2.85846	2.46537
Canada	1.77626	2.66328	2.21450	2.56175	2.35627
Chile	2.12760	2.96206	2.80484	2.67257	2.82391
China		3.11140	2.86740	3.21962	2.18632
Colombia	2.31563	2.96834	3.15352	3.13940	
Croatia	1.75666	2.96147	2.98273	3.11691	2.43001
Czech Republic	2.33676	2.78061	2.53557	2.56544	2.20023
Denmark	2.00731			2.14797	2.23760
Dominican Republic	2.75094	2.86784	3.17444	3.52405	2.99318
Egypt		2.05136	2.87055	2.77806	
El Salvador	2.87589	3.39994	3.44250	3.70209	3.20272
Estonia	1.73448	2.40385	2.44452	2.42079	2.34325
Finland	2.09136	2.58286	2.08597	2.45625	2.23000
France	2.16780	2.97815	2.21928	2.51053	2.69906
Georgia	2.41358	2.64499	3.05580	2.74663	2.41603
Germany	1.83271		2.28356	2.70390	2.18147
Great Britain	1.64677	2.54724	2.18622	2.39326	2.15212
Greece	2.12266			2.58134	2.53479
Hungary	1.73569			2.83871	2.54501
Iceland	2.17787	2.84525	2.11932	2.64113	1.85494
India	2.85364	3.58164	3.32503	3.42869	2.93711
Indonesia	2.12075	2.32837	2.63504	3.10194	
Iran	2.45482	2.39829	2.61761	2.69548	
Ireland	1.66983	2.74860	2.27580	2.54011	2.28960
Israel	2.74124	2.31262		1.99117	
Italy	2.20151	2.72860	2.21390	2.67469	2.48730
Japan	1.89048	2.20095	1.85110	2.58949	2.06766
Jordan	2.54575	2.73477	2.95101	2.81872	
Latvia	1.92223			2.66536	2.28656

Country	Left/right	Equality	Private	Government	Competition
Lithuania	2.21138	3.07459	2.89135	2.82974	2.69752
Luxembourg	1.76208	2.60595		2.46838	2.48415
Malta	1.78521			2.75868	2.08569
Mexico	2.85283	3.60650	3.30571	3.54808	3.23273
Morocco	2.93164	3.09866	3.50752	3.26359	2.21190
Netherlands	1.73625	2.02512	1.85953	2.11136	2.04367
New Zealand	1.86291	2.62809	2.26910	2.70147	2.26087
Nigeria	2.61966	2.87664		2.86610	
Norway	1.75308	2.26377	1.90572	2.34372	1.90249
Pakistan	1.46257	2.15079	1.64685	1.91684	
Peru	2.33949	2.80644	2.86111	3.22385	2.76792
Philippines	2.28830	2.71600	2.67870	2.91533	2.49010
Poland	2.29121	3.18268	2.83358	2.61729	2.76091
Portugal	2.19264		2.35647	2.72656	2.65784
Puerto Rico	2.59739	3.17348	3.10135	3.16453	2.74489
South Korea	2.21551	2.74708	2.38578	2.27107	2.23890
Macedonia	2.43438	3.02271	2.87731	2.87864	2.53609
Moldova	2.76141	2.63541	2.76377	2.69157	2.56201
Romania	2.25327	3.04189	3.23263	3.23421	2.31758
Russia	2.17428	3.00578	2.76346	2.90544	2.69169
Singapore		2.30463	2.44428	2.67522	2.06676
Slovakia	2.03019			2.61098	2.23223
Slovenia	1.82029	2.68866		2.65498	2.27948
South Africa	2.32026	3.11986	3.06017	3.11434	2.53226
Spain	1.94128	2.85794	2.49406	2.50493	2.35026
Sweden	2.08926			2.22051	1.92397
Switzerland	1.91720	3.07271	2.38122	2.71851	2.35609
Taiwan	1.69979	2.35119	2.15162	2.41842	2.03901
Turkey	2.54906	3.24765	3.29784	3.26722	3.09322
Uganda	3.14876	3.18213	3.01569	3.10382	2.31025
Ukraine	2.36851	2.98101	2.97537	2.99808	2.97123
Tanzania	3.47357	3.83358	3.83577	3.31942	3.13652
USA	1.95880	2.56661	2.23922	2.69689	2.39604
Uruguay	2.36932	3.25013	2.71755	2.86099	2.86888
Venezuela	2.65312	3.44249	3.30915	3.46030	3.01533
Vietnam	1.62831	3.06162	2.86592	2.94052	2.64291
Zimbabwe	2.51064	3.43612	3.43149	3.31763	2.62255

TABLE A2. Dependent variables

Variable name	Description and source	Obs.
Property rights index	A rating of protection of property rights on a 1 to 5 scale. The higher the score, the lower the degree of protection of property rights. Data from year 2000. Source: Miles and O'Grady (2006).	77
Business regulation index	A measure of how easy or difficult it is to open and operate a business. Scale from 1 to 5 where a higher score indicates that regulations are more burdensome for business. Data from 2000. Source: Miles and O'Grady (2006).	77
Top tax rate	Top marginal income and payroll tax rate for each country in 2000. For some countries an interval was reported and in that case an average has been used. Data from 2000, except for Georgia, Macedonia and Vietnam where data is from 2003. Source: Gwartney and Lawson (2005).	66
Corruption	An index of corruption within the political system. Scale from 1 to 6 where a higher number indicates less corruption. Average of monthly values for January 2000 to December 2000. Source: International Country Risk Guide (2001).	74
Bureaucratic delays	An index of bureaucratic delays (red tape). Scale from 0 to 4 with higher values indicating less bureaucratic delays. The data is for year 2000. Source: Business Environmental Risk Intelligence (2000).	48
Tax evasion	Assessment of the level of tax evasion. Scale from 1 to 7, where higher scores indicate less tax evasion. Data from 2001 based on surveys to executives. Source: World Economic Forum (2002)	60
Avg. government wages / GDP per capita	The ratio of average wages of central government to per capita GDP in each country. For most countries the data refers to 1995, but data for some countries are from earlier years. Source: Schiavo-Campo, de Tommaso and Mukherjee (1997).	61
Log of infant mortality	Logarithm of the number of deaths of infants per one thousand live births for the year 2000. Source: World Bank (2005).	76
Log of school attainment	Log of average number of years of schooling for adults (15+ years) for year 2000. For a few countries (Estonia, Latvia, Lithuania, Moldova and Vietnam) only data from 1990 was available and have been used. Each value is obtained as the logarithm of (1+ average years of school attainment). Source: Barro and Lee (2000).	65

Variable name	Description and source	Obs.
Literacy rate	Adult literacy rate in year 2000. For some countries, data for 2000 were not available and therefore data from the closest year in time for which data was available have been used. Data was not available for several highly developed countries and for these countries the literacy rate was set to 100 %. Source: World Bank (2005).	75
Infrastructure quality	Assessment of the quality of communication and transports. Scale from 0 to 4 with higher scores for higher quality. Data from 2000. Source: Business Environmental Risk Intelligence (2000).	48
Transfers and subsidies / GDP	Total government transfers and subsidies as a percentage of GDP. Data from 2000, except for Georgia, Macedonia and Vietnam (data from 2003). Source: Gwartney and Lawson (2005)	65
Government consumption / GDP	Government consumption expenditures as a percentage of GDP. Data from 2000, except for Georgia, Macedonia and Vietnam (data from 2003). Source: Gwartney and Lawson (2005).	72
SOEs in the economy	Index of government-owned enterprises and investments as a share of GDP. Scale from 0 to 10 where higher scores indicates a smaller share of government-owned enterprises as a share of government output. Data from 2000, except for Georgia, Macedonia and Vietnam (data from 2003). Source: Gwartney and Lawson (2005)	72
Public sector employment / total population	The ratio of public sector employment in general civilian government to total population. The data are from various years in the beginning of the 1990s. Source: Schiavo-Campo, de Tommaso and Mukherjee (1997).	61
Democracy index	Democracy index. Scale from 0 to 10, with higher values indicating a higher degree of democracy. The index is composed of an assessment of three different elements: the competitiveness of political participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive. Data from 2000. Source: Marshall & Jaggers (2002).	73
Political rights index	Index of political rights. Ranking from 1 to 7 based on an assessment of the electoral process, political pluralism and participation, and functioning of the government. Lower ratings indicate that a country comes closer to the "ideal suggested by the checklist questions". Data from 2000. Source: Freedom House (2005).	77

TABLE A3. Control variables

Variable name	Description and source	Obs.
Ethnic, religious and linguistic fractionalization	Reflects the probability that two randomly selected individuals belongs to the same ethnic, religious or linguistic group. This is calculated as one minus the sum of squared shares of each group and is therefore takes values between 0 and 1. Source: Alesina et al. (2003).	77
Legal origin	Identifies the legal origin of the Company Law or Commercial Code of each country. There are five possible origins: (1) English Common Law; (2) French Commercial Code; (3) German Commercial Code; (4) Scandinavian Commercial Code; and (5) Socialist/Communist laws. Source: La Porta et al. (1999).	78
Religion	Identifies the percentage of the population of each country that belonged to the three most widely spread religions in the world in 1980. For countries of recent formation, the data is available for 1990-95. The numbers are in percent (scale from 0 to 100). The three religions identified here are: (1) Roman Catholic; (2) Protestant; and (3) Muslim. The residual is called "other religions". Data for protestants in Lithuania was missing and has been set to 1.9 percent which is the figure reported in CIA World Fact Book 2005. Source: La Porta et al. (1999) and Central Intelligence Authority (2005).	78
Latitude	The absolute value of the latitude of the country, scaled to take values between 0 and 1. Source: La Porta et al. (1999).	78
GDP per capita (2000)	GDP per capita in US dollars (constant 2000). Source: World Bank (2005).	77
GDP per capita (1990)	GDP per capita in US dollars (constant 2000). Source: World Bank (2005).	76
GDP per capita (1980)	GDP per capita in US dollars (constant 2000). Source: World Bank (2005).	61
Trust	Average (binary) response from World Values Survey 1999-2001 to the question whether people can generally be trusted. Source: European Values Study Group and World Values Association (2004).	78
Civic Engagement	Percentage of population answering "yes" to at least one of the questions of participation in civil organizations in the WVS (a066 - a079).	56

Variable name	Description and source	Obs.
Gini index	Estimates of the Gini index based on primary household survey data obtained from government statistical agencies and World Bank country departments. Data for high-income economies are from the Luxembourg Income Study database. Data refers to various years between 1990 and 2004 and the observation closest to year 2000 have been used. Source: World Bank (2005).	74
Ethnic polarization	This meaure of polarization is highest when the population consists of two ethnic groups of equal size. From Montalvo and Reynal-Querol (2005) who use data from the World Christian Encyclopedia.	
Percent mountainous terrain	We obtain this variable from the data provided by Montalvo and Reynal-Querol (2005). They in turn base this variable on work by geographer A. J. Gerard for the World Bank's "Economics of Civil war, Crime, and Violence" project.	58
Area	Area measured in square kilometers. Data refers to year 2000. Source: World Bank (2005).	75
Total population	Data refers to year 2000. Source: World Bank (2005).	76
Population density	People per square kilometer. Data refers to year 2000. Source: World Bank (2005).	75

References

Acemoglu, D., Chernozhukov V. and Yildiz, M. (2006), "Learning and Disagreement in an Uncertain World", unpublished manuscript, Massachusetts Institute of Technology.

Alesina, A. and Angeletos, G-M. (2005), "Fairness and Redistribution: US vs. Europe", *American Economic Review* 95, 913-35.

Alesina, A., Baqir, R. and Easterly, W. (1999), "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics* 115(4), 1167–1199.

Alesina, A., Glaeser, E. and Sacerdote, B. (2001), "Why Doesn't the United States Have a European-Style Welfare State?", *Brookings Paper on Economics Activity* 2001(2), 187–254.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003), "Fractionalization", *Journal of Economic Growth* 8(2), 155–194.

Alt, J. E. and Dreyer Lassen, D. (2000), "Transparency, Political Polarization, and Political Budget Cycles in OECD Countries", *American Journal of Political Science* 50(3), 530-550.

Au, K. (1999), "Intra-Cultural Variation: Evidence and Implications for International Business", *Journal of International Business Studies* 30(4), 799–812.

Au, K. and Cheung. M. (2004), "Intra-cultural Variation and Job Autonomy in 42 Countries", *Organization Studies* 25(8), 1339–1362.

Bailey, M. and Brady, D. (1998), "Heterogeneity and Representation: The Senate and Free Trade", *American Journal of Political Science* 42(2), 524-544.

Barro, R. J. and Lee, Jong-Wha (2000), "International Data on Educational Attainment: Updates and Implications", CID Working Paper No 42, Center for International Development at Harvard University. Data retrieved from www.worldbank.org.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T. and Birch, K. (1981), "Is Empathic Emotion a Source of Altruistic Motivation?", *Journal of Personality and Social Psychology* 40(2), 290-302.

Bénabou, R. and Tirole, J. (2006), "Belief in a Just World and Redistributive Politics", *Quarterly Journal of Economics* 121, 699-746.

Biewen, M. (2000), "Income Inequality in Germany During the 1980s and 1990s", *Review of Income and Wealth* 46(1), 1-19.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995)," Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak", *Journal of the American Statistical Association* 90(430), 443-450.

Business Environmental Risk Intelligence (2000), *Historical Ratings Research Package*, Friday Harbor, WA.

Byrne, D. (1961), "Interpersonal Attraction and Attitude Similarity", *Journal of Abnormal Social Psychology* 62, 713-715.

Byrne, D. (1971), *The Attraction Paradigm*, New York: Free Press.

Central Intelligence Agency (2005), *World Fact Book 2005*, Washington, DC.

Chen, F. and Kenrick, D. (2002), "Repulsion or Attraction? Group Membership and Assumed Attitude Similarity", *Journal of Personality and Social Psychology* 83(1), 111-125.

Cukierman, A., Edwards, S., and Tabellini, G. (1992), "Seignorage and Politicak Instability", *American Economic Review* 82(3), 537-555.

Deaton, A. (2003), "Health, Inequality and Economic Development", *Journal of Economic Literature* 41(1), 113-158.

DiMaggio, P., Evans, J. and Bryson, B. (1996), "Have Americans' Social Attitudes Become More Polarized?", *American Journal of Sociology* 102(3), 690—755.

Dixit, A. and Weibull, J. (2007), "Political Polarization", *Proceedings of the National Academy of Sciences* 104, 7351-7356.
Esteban, J-M. and Ray (1994), "On the Measurement of Polarization", *Econometrica* 62(4), 819-851.

European Values Study Group and World Values Association (2004), *European and World Values Surveys Integrated Data File 1999-2002*, RELEASE I, 2nd ISCPSR version.

European Values Study Group and World Values Association (2006), *European and World Values Surveys Four-wave Integrated Data File 1981-2004, version 20060423*, downloaded from www.worldvaluessurvey.org.

Feren, D., Carroll, S. and Olian, J. (1988), "Effects of Managerial Performance and Attitudinal Similarity on Interpersonal Attraction", *Basic and Applied Social Psychology* 9(1), 33-44.

Fernández, R. and Levy, G. (2007), "Diversity and Redistribution", NBER Working Paper 11570.

Fong, C. (2001), "Social Preferences, Self-Interest, and the Demand for Redistribution", *Journal of Public Economics* 82(2), 225-246.

Freedom House (2005), *Freedom in the World Country Rankings 1973-2005*, Washington, DC. Data retrieved from www.freedomhouse.org.

Frye, T. (2002), "The Perils of Polarization: Economic Performance in the Postcommunist World", *World Politics* 54(3), 308-337.

Fukuyama, F. (1995), *Trust: The social virtues and the creation of prosperity*, Free Press, New York.

Gerber, E. and Lewis, J. (2004), "Beyond the Median: Voter Preferences, District Heterogeneity, and Representation", *Journal of Political Economy* 112(6), 1364-1383.

Glaeser, E. L., Ponzetto, G. A. M., and Shapiro, J. M. (2005), "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values", *Quarterly Journal of Economics* 120(4), p 1283-1330.

Gradstein, M. and Justman, M. (2002), "Education, Social Cohesion, and Economic Growth", *American Economic Review*, 92(4), 1192-1204.

Goff, B. and Grier, K. (1993), "On the (Mis)measurement of Legislator Ideology and Shirking", *Public Choice* 76, 5-20.

Guiso, L., Sapienza, P., Zingales, L., "Does Culture Affect Economic Outcomes?", *Journal of Economic Perspectives* 20(2), 23-48.

Gwartney, J. and Lawson, R. (2005), *Economic Freedom of the World: 2005 Annual Report*, Fraser Institute, Vancouver. Data retrieved from www.freetheworld.com.

Hamilton, W. D. (1964), "The Genetic Evolution of Social Behaviour", *Journal of Theoretical Biology* 7, 1-52.

Harrison, D., Price, K. and Bell, M. (1998), "Beyond Relational Demography: Time and the Effects of Surface- and Deep-level Diversity on Work Group Cohesion", *Academy of Management Journal*, 41(1), 96-107.

Horowitz, D. (1985), *Ethnic Groups in Conflict*, Berkeley: University of California Press.

Jehn, K., Chadwick, C. and Thatcher, S. (1997), "To Agree or Not To Agree: The Effects of Value Congruence, Individual Demographic Dissimilarity, and Conflict on Workgroup Outcomes", *The International Journal of Conflict Management* Vol. 8, No. 4 (October), 287-305.

Jehn, K., Northcraft, G., and Neale, M. (1999), "Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups", *Administrative Science Quarterly*, 44, 741-763.

Jehn, K. and Mannix, E. (2001), "The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance", *Academy of Management Journal*, Vol. 44, No. 2, 238-251.

Knack, S. and Keefer, P. (1997), "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation", *Quarterly Journal of Economics* 112(4), 1251–1288.

Knack, S. and Keefer, P. (2001), "Trust and Growth", *Economic Journal*, 111(470), 295-321(27).

La Porta, R., Lopez de Silanes, F., Shleifer, A. and Vishny, R. (1997), "Trust in Large Organizations", *American Economic Review*, 87(2), 333–338.

La Porta, R., Lopez de Silanes, F., Shleifer, A. and Vishny, R. (1999), "The Quality of Government", *Journal of Law, Economics and Organization* 15(1), 222–279.

Ledyard, J. O. (1995). "Public Goods: A Survey of Experimental Research", Chapter in *Handbook of Experimental Economics*. Eds.: J. Kagel and A. E. Roth. Princeton: Princeton University Press, pp. 111-194.

Kalt, J. and Zupan, M. (1990), "The Apparent Ideological Behavior of Legislators: testing for Principal-Agent Slack in Political Institutions", *Journal of Law & Economics* 33, 103-131.

Luttmer, E. F. P. (2001), "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy* 109(3), 500–528.

Marshall, M. G. and Jaggers K. (2002), Political Regime Characteristics and Transitions 1800-2002, Polity IV Project, University of Maryland. Data retrieved from http://www.cidcm.umd.edu/inscr/polity/.

McCleary, R. M. and Barro, R. J., "Religion and Economy, *Journal of Economic Perspectives*, 20(2), 49-72.

McGrath, J. E. (1984), *Groups: Interaction and Process*, Englewood Cliffs, NJ: Prentice-Hall.

Meltzer, A. H. and Richard, S. F. (1981), "A Rational Theory of the Size of Government", *Journal of Political Economy* 89(5), 914-927.

Miles, M., Holmes, K. and O'Grady, M. A. (2006), *2006 Index of Economic Freedom*, The Heritage Foundation and Down Jones & Company, Washington, DC. Data retrieved from www.heritage.org.

Montalvo, J, G. and Reynal-Querol, M, (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review* 95(3), 796-816.

Mouw, T. and Sobel, M. E. (2001), "Culture Wars and Opinion Polarization: The Case of Abortion", *American Journal of Sociology* 106(4), 913-943.

Newcomb, T. M. (1961), *The Acquaintance Process*, New York: Holt, Rinehart & Winston.

Oetzel, J. (1998), "Explaining Individual Communication Processes in Homogeneous and Heterogeneous Groups Through Individualism-Collectivism and Self-Construal", *Human Communication Research*, 25(2), 202-224.

Olson, J. M., Vernon, P. A. and Jang, K. L. (2001), "The Heritability of Attitudes: A Study of Twins", *Journal of Personality and Social Psychology* 80(6), 845-860.

Park, J. and Schaller, M. (2005), "Does Attitude Similarity Serve As a Heuristic Cue for Kinship? Evidence of an Implicit Cognitive Association", *Evolution and Human Behavior* 26, 158-170.

Persson, T. and Tabellini, G. (2000), *Political Economics: Explaining Economic Policy*, The MIT Press, Cambridge, MA.

PRS Group (2001), *International Country Risk Guide 2001*, East Syracuse, NY. Data retrieved from www.countrydata.com.

Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, NJ.

Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon and Schuster, New York, NY.

Roemer, J. E. (1998), "Why the poor do not expropriate the rich: an old argument in new garb", *Journal of Public Economics* 70, 399–424.

Schiavo-Campo, S., de Tommaso, G. and Mukherjee, A. (1997), *An International Statistical Survey of Government Employment and Wages*, Government Employment and Pay: A Global and Regional Perspective Series, Public Sector Management and Information Technology Team, Technical Department for Europe, Central Asia, Middle East and North Africa, World Bank, Washington, DC.

Shalom, S. and Sagie, G. (2000), "Value Consensus and Importance", *Journal of Cross-Cultural Psychology* 31(4), 465–497.

Suedfeld, P., Bochner, S. and Wnek, D. (1972), "Helper-Sufferer Similarity and a Specific Request for Help: Bystander Intervention During a Peace Demonstration", *Journal of Applied Social Psychology* 2 (1), 17–23.

Svensson, J. (1998), "Investment, Property Rights and Political Instability: Theory and Evidence", *European Economic Review* 42(7), 1317-1341.

Tesser, A. (1993), "The Importance of heritability in Psychological Research: The Case of Attitudes", *Psychological Review* 100(1), 129-142.

Vigdor, J. L. (2004) "Community Composition and Collective Action: Analyzing Initial Mail Response to the 2000 Census." *Review of Economics and Statistics* 86(1), 303–312.

World Bank (2005), *World Development Indicators 2005*, Washington, DC. Data retrieved from www.worldbank.org.

World Economic Forum (2002), *Global Competitiveness Report 2001-2002*, World Economic Forum 2001 Executive Opinion Survey, Geneva, Switzerland.

EFI, The Economic Research Institute

Published in the language indicated by the title. A complete publication list can be found at www.hhs.se/efi Books and dissertations can be ordered from EFI via e-mail: EFI.Publications@hhs.se

Reports since 2003

2007

Books

Wijkström, Filip och Torbjörn Einarsson. *Analysmodell för sektorsöverskridande statistik. Fallet vård och omsorg.*

Dissertations

Alexius, Susanna. *Regelmotståndarna – om konsten att undkomma regler.*

Berg, Bengt Åke. *Volatility, Integration and Grain Bank. Studies in Harvests, Rye Prices and Institutional Development of the Parish* Magasins *in Sweden in the 18th and 19th Centuries.*

Brodin, Karolina. *Consuming the Commercial Break: An Ethnographic Study of the Potential Audiences for Television Advertising.*

Hagberg, Axel. *Bankkrishantering: Aktörer, marknad och stat.*

Hjalmarson, Hanna. *En växande marknad: Studie av nöjdheten med konsumtionsrelaterade livsområden bland unga konsumenter.*

Hjelström, Tomas. *The Closed-End Investment Company Premium Puzzle – Model development and empirical tests on Swedish and British data.*

Kraus, Kalle. *Sven, inter-organisational relationships and control – a case study of domestic care of the elderly.*

Melian, Catharina. *Progressive Open Source*

Valiente, Pablo. *Re-innovating the Existing. A study of Wireless IS Capabilities to Support Mobile Workforces.*

2006

Books

Lundeberg, Mats, Pär Mårtensson and Magnus Mähring (eds) *IT & Business Performance. A Dynamic Relationship. EFI Yearbook 2006.* EFI/Studentlitteratur.

Thodenius, Björn. *Organisering av kunskap. En studie av Wallenberg Consortium North.*

Wijkström, Filip och Torbjörn Einarsson. *Från nationalstat till näringsliv? Det civila samhällets organisationsliv i förändring.*

Wijkström, Filip, Einarsson, Stefan och Larsson, Ola. *Staten och det civila samhället. Idétraditioner och tankemodeller i den statliga bidragsgivningen till ideella organisationer.*

Östman, Lars. *Lysande ögonblick och finansiella kriser. Dramaten under ett sekel.*

Dissertations

Argenton, Cedric. *Quality provision in duopoly.*

Beckerman, Carina. *The Clinical Eye. Constructiong and Computerizing an Anesthesia Patient Record.*

Borglund, Tommy. *Aktievärden i fokus – internationell påverkan på intressentrelationer vid förvärv och fusion.*

Breman, Anna. *The Economics of altruism, paternalism and self-control.*

Edquist, Harald. *Technological breakthroughs and productivity growth.*

Eklund, Jana. *Essays on Forecasting and Bayesian Model Averaging.*

Frostenson, Magnus. *Legitimitetskontrollen – en studie av etiska värderingars roll i gränsöverskridande förvärv och fusioner.*

Gaspar, Raquel M. *Credit Risk and Forward Price Models.*

Gustafsson, Peter. *Essays on Trade and Technological Change.*

Hopkins, Elisabeth. *Is a higher degree of local currency pricing associated with lower exchange rate pass-through? A study of import pricing in 51 Swedish industries.*

Kling, Ragnar. *Developing Product Development in Times of Brutal Change.*

Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*

Lychnell, Lars-Olof. *"Och fungerar det inte, gör vi på något annat sätt" – En klinisk fallstudie av IT-relaterat förändringsarbete i småföretag*

Meitz, Mika. *Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data.*

Mendicino, Caterina. *Financial Market Imperfections, Business Cycle Fluctuations and Economic Growth.*

Ovanfors, Anna. *Essays on Nonlinear Time Series Analysis and Health Economics.*

Paltseva, Elena. *Essays on Commitment and Inefficiency in Political Economy.*

Rogberg, Martin. *Den modeföljande organisationen. Om acceptansen av TQM och andra populära managementmodeller.*

Silvennoinen, Annastiina. *Essays on Autoregressive Conditional Heteroskedasticity.*

Sjögren, Ebba. *Reasonable Drugs. Making Decisions with Ambiguous Knowledge.*

Slinko, Irina. *Essays in Option Pricing and Interest Rate Models.*

Wilander, Fredrik. *Essays on Exchange Rates and Prices.*

2005

Books

Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks – boundaries and new connections.*

Charpentier, Claes. *IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.*

Dembrower, Maria. *Entreprenörskap i industriella nätverk.*

Lind, Johnny och Göran Nilsson (red). *Ekonomistyrningens metoder, sammanhang och utveckling – En vänbok till Lars A Samuelson.*

Samuelson, Lars A. *Organizational governance and control – a summary of research in the Swedish society.*

Svedberg Nilsson, Karin, Roger Henning och Karin Fernler (eds). *En illusion av frihet? Företag och organisationer i regelsamhället. EFIs Årsbok 2005.* EFI/Studentlitteratur.

Dissertations

Andersson, Martin. *Making a Difference – Project Result Improvement in Organizations.*

Arvidsson, Per. *Styrning med belöningssystem – Två fallstudier om effekter av belöningssystem som styrmedel.*

Bems, Rudolfs. *Essays in International Macroeconomics.*

Berg-Suurwee, Ulrika. *Nya trender, nya nämnder – effekter av en stadsdelsnämndsreform inom kultur och fritid.*

Björkman, Hans. *Learning from members – Tools for strategic positioning and service innovation in trade unions.*

Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice.*

Clapham, Eric. *Essays in Real Estate Finance.*

Dareblom, Jeanette. *Prat, politik och praktik – Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.*

Fromm, Jana. *Risk Denial and Neglect: Studies in Risk Perception.*

Hjelström, Anja. *Understanding International Accounting Standard Setting – A Case Study of IAS 12, Accounting for Deferred Tax.*

Hortlund, Per. *Studies on Swedish Banking 1870-2001.*

Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas.*

Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures.*

Madestam, Andreas. *Developing Credit Markets.*

Nilsson, Roland. *The Market Impact of Short-Sale Constraints.*

Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods.*

Nordin, Fredrik. *Externalising Services – Walking a Tightrope between Industrial and Service Logics.*

Parmler, Johan. *Essays in Empirical Asset Pricing.*

Simbanegavi, Witness. *Price Discrimination, Advertising and Competition.*

Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag.*

Tolis, Christofer. *Framing the Business – Business Modelling for Business Development.*

Östberg, Per. *Corporate Disclosure and Investor Recognition.*

2004

Books

Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen.*

Lind, Johnny. *Strategi och ekonomistyrning. En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning.*

Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman.*

Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet. EFIs Årsbok 2004.*

Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden – Their scope, roles and visions.*

Dissertations

Anderson, Anders. *Essays in Behavioral Finance.*

Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists.*

Blix, Magnus. *Essays in Mathematical Finance – Modelling the Futures Price.*

González Gómez, Andrés. *Nonlinear dynamics and smooth transition models.*

Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.*

Ivarsson Westerberg, Anders. *Papperspolisen – varför ökar administrationen i moderna organisationer.*

Jutterström, Mats. *Att påverka beslut – företag i EUs regelsättande.*

Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition.*

Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.*

Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices.*

Malmsten, Hans. *Properties and Evaluation of Volatility Models.*

Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments.*

Mattsson, Susanna. *På gränsen mellan ordning och oordning – tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*

Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*

Nilsson, Hans. *Medborgaren i styrsystemet – beskrivning av VAD och HUR i styrning av kommunal verksamhet.*

Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*

Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*

Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*

Salabasis, Mickael. *Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.*

Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*

Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*

Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*

Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*

Talia, Krim. *The Scandinavian Currency Union, 1873–1924 – Studies in Monetary Integration and Disintegration.*

2003

Books

Lundahl, Mats (ed). *Globalization and Its Enemies. EFIs Årsbok 2003.*

Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (eds). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*

Bergman, Malin. *Essays on Human Capital and Wage Formation.*

Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*

Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*

Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*

Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*

Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*

Lange, Fredrik. *Brand Choice in Goal-derived Categories – What are the Determinants?*

Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*

Magnusson, Peter R. *Customer-Oriented Product Development – Experiments Involving Users in Service Innovation.*

Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*

Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*

Nilsson, Göran. *Processorientering och styrning – Regler, mål eller värderingar?*

Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*

Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*

Tillberg, Ulrika. *Ledarskap och samarbete – En jämförande fallstudie i tre skolor.*

Waldenström, Daniel. *Essays in Historical Finance.*

Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

Ögren, Anders. *Empirical Studies in Money, Credit and Banking – The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834–1913.*

Essays on Privatization, Identity, and Political Polarization

Erik Lindqvist

AKADEMISK AVHANDLING

Som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
torsdagen den 18 oktober 2007, kl. 13.15 i sal 750,
Handelshögskolan, Sveavägen 65