Three Essays in Applied Regulation

Andres Gomez-Lobo

September 1997

A Thesis Submitted for the Degree of Doctor of Philosophy to University College London, the University of London.
Abstract

In the last fifteen years, theoretical economists have reformulated the theory of regulation using the tools of informational economics. Although theoretical advances in the field have been numerous, empirical work on the subject has been sparse. The main objective of this thesis is to apply empirical techniques to different topics in the area of regulation. In agreement with the new theory, in each of the essays special emphasis is placed on informational issues.

The first essay concerns demand side aspects of regulation. Recent regulatory reforms in the United Kingdom energy markets will result in important changes in domestic energy tariffs. A demand system for the United Kingdom is estimated using data from the Family Expenditure Survey. The estimation strategy takes into account problems raised by the non-linear nature of energy prices and unobserved consumer heterogeneity. The results are subsequently used to calculate the welfare consequences of changes in domestic gas tariffs. The results show that tariff rebalancing, sparked by the introduction of competition for the supply of gas, will most probably pose distributional conflicts. The results are not only useful for the design of welfare policy but also for the discussions surrounding regulatory reform.

The next two essays deal with the supply side of regulation. The first develops a structural econometric method for estimating the parameters of a cost function under conditions of asymmetric information. The method is applied to the Norwegian Bus industry. Using the estimated model, the effects of changes in the regulatory environment can be analysed. In particular, the effects of introducing an optimal second-best contract or several
types of auctions are examined. To the best of our knowledge, this is the first time in the applied regulation literature that an optimal second best contract has been characterised using information recovered exclusively by econometric methods.

The final essay deals with auctions. It examines the tendering of refuse collection contracts by local authorities in England. A structural econometric model is estimated to recover firms' underlying cost distribution. The results are used to answer several questions regarding these auctions, in particular whether there is a bias favouring one type of firm in the tendering process. This study adds one more example to the emerging econometrics of auctions literature.
# Contents

Abstract \hspace{2cm} 1

Acknowledgements \hspace{2cm} 11

Declaration \hspace{2cm} 13

Abbreviations \hspace{2cm} 15

1 Introduction \hspace{2cm} 17

1.1 Background \hspace{2cm} 17

1.2 The regulatory model \hspace{2cm} 19

1.3 The optimal regulatory scheme \hspace{2cm} 23

1.3.1 Symmetric information \hspace{2cm} 24

1.3.2 Asymmetric information \hspace{2cm} 24

1.3.3 Remarks \hspace{2cm} 30

1.3.4 Three essays in applied regulation \hspace{2cm} 31

1.4 The resource allocation side of regulation \hspace{2cm} 31

1.4.1 Linear prices with homogenous consumers \hspace{2cm} 33

1.4.2 Two-part tariffs with homogenous consumers \hspace{2cm} 34

1.4.3 Two-part tariffs with heterogeneous consumers \hspace{2cm} 36

1.4.4 Non-linear pricing \hspace{2cm} 39

1.4.5 Pricing and regulation schemes \hspace{2cm} 41

1.4.6 Empirical applications and the first essay \hspace{2cm} 42

1.5 The supply side of regulation \hspace{2cm} 45
1.5.1 Implementation and menus of linear contracts ........................................ 46
1.5.2 Price-cap regulation, rate of return regulation and profit sharing ........... 48
1.5.3 Applicability of the optimal contract .................................................. 50
1.5.4 Review of the empirical literature on regulation .................................... 53
1.5.5 The Norwegian bus transport study .................................................. 61

1.6 Rent extraction and auctions in regulation ............................................. 63
1.6.1 Auctioning incentive contracts .......................................................... 64
1.6.2 Non-structural empirical studies of auctions ....................................... 67
1.6.3 Structural econometric models of auctions ......................................... 71
1.6.4 Asymmetric auctions and the tendering of refuse collection contracts in England .......................................................... 86

1.7 Conclusions ............................................................................................ 88

I Regulation: Demand issues ........................................................................ 91

2 Energy Demand Modelling and Tariff Rebalancing in the United Kingdom Gas Market ................................................................. 93
2.1 Introduction .......................................................................................... 93
2.2 Tariffs, price discrimination and the tax benefit system ..................... 97
  2.2.1 Optimal Tariffs ................................................................................ 99
  2.2.2 Tariffs and the tax-benefit system .................................................. 102
2.3 Household energy demand model ......................................................... 104
  2.3.1 The QUAIDS demand system ....................................................... 104
  2.3.2 Two-part tariffs and econometric estimation .................................. 107
  2.3.3 Data ............................................................................................. 115
  2.3.4 Estimation results ........................................................................ 118
2.4 Welfare Calculations ............................................................................ 126
2.5 Conclusions .......................................................................................... 135
Appendix 2.A ............................................................................................... 137
II Regulation: Supply issues

3 Empirical study of the Norwegian bus industry

3.1 Introduction ......................................................... 141
3.2 Incentives in regulation ........................................... 142
3.3 Regulatory practice in the Norwegian bus industry ...... 148
3.4 The structural model ................................................ 151
    3.4.1 Data ............................................................ 158
    3.4.2 Stochastic specification .................................... 159
    3.4.3 Estimation and Results ..................................... 160
3.5 Welfare calculations .............................................. 163
3.6 Conclusions ......................................................... 173
Appendix 3.A ............................................................ 175

III Regulation: Auctions

4 Tendering of refuse collection contracts in England

4.1 Introduction ......................................................... 181
4.2 History of CCT ...................................................... 183
4.3 Data and descriptive statistics ................................. 185
4.4 Preliminary data analysis: do the number of bids matter? 189
    4.4.1 Auction theory ................................................. 189
    4.4.2 Empirical specification ..................................... 193
    4.4.3 Results .......................................................... 196
    4.4.4 Conclusions of preliminary analysis ...................... 207
4.5 Structural analysis ................................................ 209
4.6 Asymmetric private value auction .............................. 214
4.7 Estimation method ............................................... 216
4.8 Specification, estimation and results .......................... 220
    4.8.1 Comparing the estimated cost distributions .......... 226
IV Conclusions

5 Final Remarks

References
List of Tables

2.1 Probit Estimation of Gas Connection Equation .................. 119
2.2 Gas and electricity share equations with symmetry imposed . 123
2.3 Weighted average of individual elasticities ....................... 125
2.4 Welfare Effects of Tariff Rebalancing ............................. 128
2.5 Average compensating variation for different groups ........... 131
2.6 Net welfare effects of tariff rebalancing (case 1) combined with
cost reductions ................................................................. 134
2.7 Food, Clothing and Alcohol share equations with symmetry
imposed ............................................................................ 138

3.1 Structural cost function estimation results ......................... 162
3.2 Results for different contracts and different costs of public
funds at sample average ...................................................... 167
3.3 Rents, transfers and welfare of an optimal auction ............... 171
3.4 Rents, transfers and welfare of a Vickrey auction of fixed price
contracts ................................................................. 172
3.5 Summary statistics of data for the Norwegian bus transport
industry ................................................................. 175
3.6 Estimation results with alpha 10% higher ......................... 176
3.7 Estimation results with alpha 10% lower ......................... 177

4.1 Bidding and expenditure ................................................. 186
4.2 Differences related to political control ......................... 188
4.3 Expenditure on refuse collection regressions .................. 197
4.4 Probability of DSO winning contract ................................................................. 199
4.5 Cost reduction effects of the number of bids .................................................. 200
4.6 Regressions of first differenced model ............................................................... 203
4.7 Hausman specification tests .................................................................................. 206
4.8 Probit estimation of DSO success rate with endogeneity correction ...................... 208
4.9 Cost reduction effects of the number of bids implied by the first differenced model. ................................................................. 209
4.10 Binomial test for DSO success rate .................................................................. 210
4.11 Summary statistics of cost pseudo-samples ...................................................... 225
4.12 Test results for distribution differences ................................................................ 227
4.13 Test for distribution differences (trimmed valuations) ......................................... 229
4.14 DSO success rate by semester ........................................................................... 232
4.15 Breakpoints and observations for categories used in the equality of distributions test ............................................................................................. 235
List of Figures

2.1 Expenditure shares of gas and electricity. .......................... 107
2.2 Budget constraint with two-part tariff ............................... 108
2.3 Corner solution with two-part tariff ................................. 109
2.4 Compensating variation and total expenditure (case 1) ......... 130
2.5 Frequency distribution of compensating variation (case 1). ... 132

3.1 Cost ratio vs Output ratio for 1991 cross section (solid line
above 45 degree line is predicted cost ratio from OLS regression). 160
3.2 Empirical distribution of theta ........................................ 165
3.3 Fitted normal distribution .............................................. 166
3.4 Effort level as a function of type, cost of public funds = 0.3. 168
3.5 Cost as a function of type, cost of public funds = 0.3. ......... 169

4.1 Pseudo costs and bids ................................................... 222
4.2 Distribution and density of costs for DSO. ......................... 224
4.3 Distribution and density of costs for private firms ............... 225
4.4 Cumulative distributions for cost valuations ...................... 230
Acknowledgements

There is a long list of people and institutions that have supported me during the course of my doctoral studies. To all of them I would like to express my most sincere gratitude.

For financial support I would like to thank University College London for their Open Graduate Scholarship. Also, the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom (CVCP) for their Overseas Research Students Award (ORS), and again to University College London for supporting my ORS application. To the Department of Economics, at University College London, for the Teaching Assistantship appointments for several years. Foremost, I am deeply indebted to the Institute for Fiscal Studies for their Graduate Scholar Award. To this last institution I am grateful for much more than financial support. It provided marvellous facilities, a stimulating academic environment, and a very friendly atmosphere in which to do research.

My ten month visit to QREMAQ, Toulouse, France, was made possible thanks to the ENTER program of which UCL is a member. I thank Jaques Cremer, Tilman Borgers and Christian Dustman for helping to arrange the visit. I would like to express my special thanks to Marc Ivaldi, for his advice and supervision when I was in Toulouse.

None of the following pages would have been possible without the support, encouragement and general supervision of Costas Meghir and Richard Blundell. Their patience with my numerous, and often trivial, questions was remarkable. There are many other people at the IFS that I would like to
thank, especially Andrew Dilnot and Robert Markless. I am also indebted to James Banks and Ian Crawford for their invaluable help on many occasions and their patience with my constant interruptions. Early discussion with Phil Burns were also very useful. I would also like to thank Dag Morten Dalen for the many discussions regarding our joint work. My knowledge of CCT, and parts of the data used in chapter 4, is in large part due to the generous help of Stefan Szymanski.

Alan Turnbull and Sheela Vyas of the Department of the Environment were very helpful in providing data for the final chapter. Desmond Yew’s help in making the bidding dataset used in the final chapter was crucial. Material from the Family Expenditure Survey, made available by the Central Statistical Office through the ESRC Data Archive, has been used by permission of the Controller of HMSO. Neither the CSO nor the ESRC Data Archive bears responsibility for the analysis or the interpretation of the data reported here.

An intense endeavour such as a PhD would be considerably more difficult without the support of friends and family. I would like to thank Martina, Philippe, João, Mary (and of course Dolly!), Jaime and Monica, Max and Emily, my parents Alfonso and Jimena, and the numerous other people that one way or another have contributed indirectly to the completion of this thesis.
Declaration

1. No part of this thesis has been presented to any University for any degree.

2. Chapter 3 was undertaken as joint work with Dag Morten Dalen.

Andrés Gómez-Lobo
Abbreviations

CCT Compulsory Competitive Tendering
FES Family Expenditure Survey
OFGAS Office of Gas Supply
Chapter 1

Introduction

1.1 Background

Regulation can be defined as the industrial organisation side of public economics. It deals with a wide range of issues that arise at the interface between markets and the state. In the present study, the focus will be on the regulation of natural monopolies. This topic has been of growing interest in the last decade, both to economist as well as policy makers.

From a political perspective, the interest in regulation has grown as the end of the cold war has changed the nature of economic debate between people from differing ideological camps. No longer are energies spent in protracted discussions about the advantages or disadvantages of global economic systems (socialism versus capitalism). Now that there is a relative consensus regarding the advantages of a mixed market economy, debate focuses on the narrower topic of the size of the State, the privatisation or not of public assets, or the need for government intervention in the functioning of certain markets. Regulatory economics is the analytical scrutiny of the case for State intervention and the optimal way to achieve this intervention. Therefore, it deals with issues that are in the forefront of current political debate.

Partly as a consequence of the above developments, in the past decade many countries, including developing nations, have embarked on diverse lib-
eralisation and privatisation programs. In most cases these reforms have heightened the role of markets and private firms in activities traditionally reserved for public entities. Special among these experiences is the privatisation of public utilities in the United Kingdom. These reforms created a pressing need to answer many questions concerning the optimal regulation of private natural monopolies.

The growing practical interest in regulation has been paralleled by the expanding economic literature in this field. The publication, in 1982, of the seminal work by Baron and Meyerson, marked the beginning of what has been called the "New Theory of Regulation" (Laffont (1994)). Unlike the received literature, the new theory of regulation reformulated the regulatory problem using the tools of informational economics. This methodological change introduced the notion of asymmetric information between regulators and firms, the idea that firms may have some information that the regulator cannot observe. This information may be crucial for the design of regulatory policy, but firms will not in general have the incentive to reveal it to the regulator. Using tools of mechanism design and agency theory, the new economics of regulation has enabled researchers to characterise the optimal regulation in such a situation. Furthermore, the analytical framework has provided a much deeper understanding of the welfare properties of different regulatory schemes and, therefore, their relative merits.

Although the theoretical literature on regulation has made great strides in the last decade, empirical work on the subject has been sparse. In part this is due to the general difficulties encountered with the empirical implementation of game theoretic models (Bresnahan, 1996). In these models, it is often the case that key variables are not easily measured and observed empirically. The concept of a "threat", so important in enforcing certain equilibrium behaviour, is a case in point. Sometimes models give multiple equilibrium solutions, without the researcher knowing which is the one which gave rise to the data in question. In addition, counter examples to a par-
ticular theoretical result abound. The difficulty of obtaining adequate data sets have also limited research in this area, as well as purely computational difficulties related to the complexity of the models in question.

However, some progress in empirical industrial organisation has been achieved\(^1\). The main objective of this thesis is to apply empirical techniques to different topics in the area of regulation. Hopefully these essays are a contribution to the bridge building that is required between theoretical developments in the field of regulation and its empirical application.

In this chapter we will review the empirical literature on regulation, describe how the three essays of this thesis fit into this literature, and explain how they expand the current knowledge in the field.

Before going into more detail about the specific essays presented below, it is convenient to develop a canonical model of regulation that serves to clarify and unify the ensuing discussion. The model presented below relies heavily on Laffont and Tirole (1993).

### 1.2 The regulatory model

We assume a market for goods or services which can only be served efficiently by one supplier. The justification for the one supplier constraint is that economies of scale, or scope, are strong enough to render the activity a natural monopoly.

The energy industries, such as gas and electricity, are prime examples of traditional natural monopolies. Not all sections of these industries are natural monopolies however. For example, scale economies may not be so severe in the extraction of natural gas or the generation of electricity. In these activities competition may be feasible and desirable. Efficiency losses due to untapped scale economies, or the duplication of fixed assets, may be more than compensated by the benefits of competition. This observation

\(^1\)See Bresnahan (1996) for a survey and discussion of empirical IO.
has informed many recent reforms in electricity industries around the world. Several countries have, by now, introduced competition and market forces in the generation side of the electricity market. Current wisdom is that only transportation and, arguably, final distribution to consumers are natural monopoly activities. The model presented below, then, would apply to these last activities.

We assume that the cost structure of the industry is

$$C = C(\theta, e, q)$$

(1.1)

where $\theta$ represents an intrinsic efficiency parameter of the firm, $e$ is the amount of cost reducing effort applied by managers and $q = (q_1, ..., q_n)$ is a vector of outputs. Higher levels of effort reduce costs for a given efficiency level and output vector. The opposite is true for $\theta$. Conditional on effort and production, a higher $\theta$ implies higher costs. These relationships are formalised as

$$\frac{\partial C}{\partial \theta}(\theta, e, q) > 0$$
$$\frac{\partial C}{\partial e}(\theta, e, q) < 0$$
$$\frac{\partial C}{\partial q_i}(\theta, e, q) > 0.$$

The objective function of the firm is,

$$U = S + R(q) - C(\theta, e, q) - \psi(e)$$

(1.2)

where $R(q)$ is revenue, $\psi$ is the disutility of effort of the firm’s managers, and $S$ is a subsidy (which can be positive or negative) provided by the regulator to the firm. Since $S$ is set by the regulator, we can assume, without loss of

$^2$The $\psi$ function is assumed to posses the following characteristics, $\psi' > 0$, $\psi'' > 0$, and $\psi''' \geq 0$.  

20
generality, the accounting convention that the regulator receives $R$ and pays the firm’s costs, $C$, and only gives a net transfer $t$ to the firm,

$$U = t - \psi(e).$$ (1.3)

Notice that (1.2) assumes that the firm is risk neutral. In this simple model both the firm and the regulator are assumed to be risk neutral.

The informational structure is the following. The firm knows $\theta$ and it chooses $e$. The regulator, however, does not observe either. He only observes overall ex-post costs, $C$, and has an imprecise knowledge of the possible values of $\theta$. This last idea is formalised by assuming that the regulator’s beliefs can be represented by a distribution over $\theta$, $F(\theta)$ on $\theta \in [\theta, \bar{\theta}]$. $F(\theta)$ has a continuous and positive density $f(\theta)$ on the interval $[\theta, \bar{\theta}]$. In contrast to efficiency and effort, the regulator is assumed to have full knowledge of the cost of effort function $\psi$. This may seem arbitrary. However, some limited form of uncertainty about the function $\psi$ may be possible, as long as a suitable transformation of variables allows the problem to be recast in the above structure, where the regulator only has asymmetric information regarding the cost function.$^3$

The demand side will be left purposely vague for the moment. We will assume that output, $q$, provides a social surplus of $V(q)$. We will discuss further below what $V(q)$ is in each of several alternative situations.

The final ingredient is $\lambda$, the social cost of public funds. In the present model we will interpret this parameter as the social cost of distortionary taxation needed to raise public funds. If we assume that the transfers to and from the industry being regulated are small relative to total tax revenues in the economy, it is reasonable to assume that $\lambda$ is constant.$^4$

---

$^3$See Laffont and Tirole (1993) page 36. Other regulatory models assume that costs are known but the regulator has asymmetric information regarding the disutility of effort function.

$^4$If the regulator is not allowed to make transfers to the firm, then $\lambda$ would no longer be constant. In this case any transfer given to the firm has to be funded by the firm’s
of introducing the social cost of public funds is to model the fact that most regulators want to limit the rents earned by regulated firms. If there is a social cost to transfers, then a welfare maximising regulator will attempt to limit rents.

Another reason for limiting rents is distributive. In this case the welfare of consumers has a higher weight in the regulator's objective function than the firm's profits. Taking this approach does not change the model significantly. A parameter analogous to \( \lambda \) is introduced in the welfare function to reflect the relative weight given to consumer surplus vis-à-vis profits. Baron and Meyerson (1982) cast their original model in this form.

Both the cost of public funds approach and the distributive approach lead to qualitatively similar optimal contracts. However, it is important to recognise that both approaches penalise profits in the objective function. There is a very important reason for wanting the model to have this characteristic. If profits and consumer welfare have the same welfare weight, or there is no cost of public funds, then the first best (symmetric information) outcome can be implemented even when a regulator is faced with asymmetric information regarding production costs (Loeb and Magat (1979)). In this case the regulator could give a transfer to the firm equal to net consumer surplus. Firms would then have an incentive to expand output to the socially optimal level. Needless to say, this is not a very popular regulatory mechanism. Introducing a cost of public funds is a way to model the fact that (benevolent) regulators try to minimise the rents they leave to firms.

A utilitarian regulator would try to maximise expected social welfare revenues, which requires price distortions. In this case, the social cost of transfers will depend on the level of transfers and thus \( \lambda \) will be endogenous.

5See also Armstrong, Cowan and Vickers (1994) for an exposition of regulatory theory using a distributional parameter in the objective function.

6In fact, except for the multipliers the formulas are identical.

7To be fair, Loeb and Magat (1979) propose auctioning the right to be the monopolist as an instrument to transfer the surplus back from firms to consumers. We postpone the discussion of auctions in regulation until later.
given by,

\[ E[W] = \int_\theta [V(q) - (1 + \lambda)(C(\theta, e, q) + t) + U] f(\theta) d\theta. \tag{1.4} \]

Using equation (3.11) this can be expressed as,

\[ E[W] = \int_\theta [V(q) - (1 + \lambda)(C(\theta, e, q) + \psi(e)) - \lambda U] f(\theta) d\theta. \tag{1.5} \]

The objective function (1.5) forms the basis of the forthcoming analysis. The maximisation of this objective function under symmetric and asymmetric information will reveal much about the nature of regulation and the structure of the optimal regulatory mechanism.

### 1.3 The optimal regulatory scheme

The maximisation of (1.5) is a problem in mechanism design. Although there is an adverse selection parameter, \( \theta \), and a moral hazard variable, \( e \), the model is technically an adverse selection problem. This follows from the assumption that the regulator observes costs. To see this, we can define

\[ E = E(\theta, C, q) \tag{1.6} \]

\( E \) to be the level of effort required by a firm of efficiency \( \theta \), to produce \( q \) at the overall cost of \( C \). By substituting (1.6) into the cost function and the disutility of effort function we can express the problem as a function of only one unobservable, the adverse selection parameter \( \theta \). Standard techniques in mechanism design are used to obtain the optimal regulatory contract.\(^8\)

\(^8\)Notice that some assumptions must be made regarding consumer preferences to justify the above welfare function. In particular, as discussed in Baron and Meyerson (1982), consumers must be risk neutral for expected surplus to maximise welfare. In addition, consumers must have additively separable utility between the vector of products and other expenditure.

\(^9\)The Baron and Meyerson (1982) model differs from the one presented above in that costs are not observable.

\(^10\)Murphy (1971) is the seminal work in the field of mechanism design. Fundamental and Tirole (1991), chapter 7, present a very useful review of the techniques and assumptions involved.
1.3.1 Symmetric information

It is useful to contrast the perfect information case with the asymmetric information case. Therefore, we will first examine what the optimal regulatory contract would be in the hypothetical case that the regulator knew the intrinsic efficiency of the firm, \( \theta \). In this case the regulator would instruct the firm to exert a level of effort and produce output \( q \), conditional on \( \theta \), so as to maximise the integrand of (1.5). The first order conditions for this maximisation would be,

\[
\frac{\partial V}{\partial q_i} = (1 + \lambda) \frac{\partial C}{\partial q_i}(\theta, e^*, q^*) \tag{1.7}
\]

\[
\frac{\partial \psi}{\partial e}(e^*) = - \frac{\partial C}{\partial e}(\theta, e^*, q^*) \tag{1.8}
\]

\[
t = \psi(e^*). \tag{1.9}
\]

Condition (1.7) simply states that output should be such that the marginal social value is equal to the marginal social cost. This is the traditional welfare maximising output condition from microeconomics, except for a mark-up due to the cost of public funds. The social cost of production is higher than private costs by a factor of \( 1 + \lambda \).

The second condition states that effort should be exerted until the social marginal benefit, in terms of reduced costs, equals the marginal disutility of effort. The third condition states that transfers should be such that utility is zero for all types of firms. Due to the cost of public funds, rents left to the firm are costly and the regulator, in this ideal situation of perfect information, would set them equal to zero.

1.3.2 Asymmetric information

When the regulator does not observe the efficiency of the firm, the optimal regulatory contract is found by appealing to the revelation principle. This
states that if there is an optimum strategy (as a function of efficiency type \( \theta \)) for the firm facing a particular regulatory contract, then this contract can also be implemented as a direct truth revealing mechanism. To be more precise, in the model presented above there are \( n + 2 \) observable variables: \( n \) outputs, \( C \) and \( t \). Faced with a particular regulatory contract, a firm with efficiency \( \theta \) will find it optimal to produce a certain amount of output, at a certain cost and receive transfers, say \( \{ q^*(\theta), C^*(\theta), t^*(\theta) \} \). The regulator could just as well have asked the firm to announce an efficiency parameter, \( \hat{\theta} \), and then commit to implement the policy \( \{ q^*(\hat{\theta}), C^*(\hat{\theta}), t^*(\hat{\theta}) \} \). Since the outcome for the firm from this direct contract is exactly equal to what the company would have received in the indirect case, it is in the firm’s interest to announce its true type. Thus the optimal announcement is \( \hat{\theta} = \theta \).

Since the equilibrium of any indirect mechanism can be implemented by a direct truth revealing mechanism, we can concentrate on finding the optimal contract within the class of direct truth revealing mechanisms.

Denote \( U(\hat{\theta}, \theta) \) the utility of a firm that announces \( \hat{\theta} \) when its true type is \( \theta \). A truth revealing mechanism is a set \( \{ q(\hat{\theta}), C(\hat{\theta}), t(\hat{\theta}) \} \), as a function of the announcement \( \hat{\theta} \), such that,

\[
U(\theta; \theta) \geq U(\hat{\theta}; \theta) \quad \forall \, \hat{\theta}, \theta \in [\bar{\theta}, \bar{\theta}],
\]

\[
U(\theta; \theta) \geq 0 \quad \forall \, \theta \in [\bar{\theta}, \bar{\theta}].
\]

The firm announces an efficiency type \( \hat{\theta} \), and the regulator commands the firm to produce \( q(\hat{\theta}) \), at a cost of \( C(\hat{\theta}) \), and gives the firm a transfer equal to \( t(\hat{\theta}) \). The incentive compatibility (IC) constraint states that, for the mechanism to be truth revealing, the utility of announcing the true efficiency is higher than the utility from announcing a non-true type. Thus, a truth revealing mechanism has to be designed so that firms do not have the incentive to pretend to be more or less efficient than what they truly are.
The individual rationality (IR) constraint states that all firms obtain non-negative utility. Otherwise, some types would refuse the contract altogether and would not participate in the regulatory game.

Locally, the incentive compatibility constraint will hold if a firm has no incentive to pretend to be slightly less efficient than it really is. Looking at the firm’s utility, and substituting equation (1.6), we have,

$$U(\hat{\theta}; \theta) = t(\hat{\theta}) - \psi(E(\theta, C(\hat{\theta}), q(\hat{\theta}))).$$

(1.10)

The firm will make an announcement so as to maximise (3.14). In order for the (IC) constraint to hold, the first order condition, evaluated at an announcement equal to true efficiency, must be zero. Formally,

$$\frac{\partial U}{\partial \theta}(\theta; \theta) = t'(\theta) - \psi'(E(\theta)) \left[ \frac{\partial E}{\partial C} \frac{\partial C}{\partial \theta} + \sum_{i=1}^{n} \frac{\partial E}{\partial q_i} \frac{\partial q_i}{\partial \theta} \right]_{\hat{\theta} = \theta} = 0.$$

By the envelope theorem, this implies that the utility of firms must evolve according to,

$$\frac{dU}{d\theta} = -\psi'(E(\theta, C(\theta), q(\theta))) \frac{\partial E}{\partial \theta} < 0.$$  

(1.11)

Equation (1.11) states that for the mechanism to be truth revealing it must give higher utility (rents) to efficient types (low \(\theta\)) than to inefficiency types (high \(\theta\)). The intuition is the following: an efficient firm could pretend to be slightly less efficient, and it would therefore have to exert less effort to reach the same cost and output vector expected of the less efficient firm. This effort savings is \(\frac{\partial E}{\partial \theta}\) and the marginal cost of effort is \(\psi'(e)\). Therefore, a firm of type \(\theta\) could obtain a utility of \(\psi'(\theta)\frac{\partial E}{\partial \theta}\) by declaring a slightly higher efficiency parameter. In order for the mechanism to give incentives for firms to reveal their true type, it must give them at least this utility for declaring their true type. Equation (1.11), required by incentive compatibility, is the source of the socially costly informational rents that a regulator must give firms under asymmetric information.
Equation (1.11) is necessary but not sufficient for the (global) condition (IC) to hold. The local condition (1.11) is sufficient if, in addition, the mechanism is such that,

\[
\frac{dC}{d\theta} \geq 0 \tag{1.12}
\]
\[
\frac{dq_i}{d\theta} \leq 0 \ \forall \ i. \tag{1.13}
\]

These second order conditions imply that the mechanism must compel less efficient firms to produce less and also have higher costs\(^{11}\).

Because incentive compatibility requires more efficient firms to have higher utility, the individual rationality constraint (IR) will only be binding for the most inefficient type (\(\bar{\theta}\)). Since leaving rents to firms is welfare reducing, we are assured that the optimal contract will set utility as low as possible. Therefore, we can simplify the (IR) constraint to \(U(\bar{\theta}) = 0\).

We have characterised the conditions that a truth revealing mechanism must meet, and by the revelation principle we know that a regulatory contract can be implemented by a direct truth revealing mechanism. Therefore, to find the optimal regulatory contract we need to find the direct mechanism that maximises expected social welfare (1.5). That is, the optimal contract will,

\[
\max_{q(\cdot),e(\cdot),U(\cdot)} \int_{\bar{\theta}} [V(q) - (1 + \lambda) (C(\theta, e, q) + \psi(e)) - \lambda U] f(\theta)d\theta \tag{1.14}
\]

subject to,

\[
\frac{dU}{d\theta} = -\psi'(e) \frac{\partial E}{\partial \theta} (\theta, C(\theta, e, q), q) \tag{IC}
\]
\[
U(\bar{\theta}) = 0 \quad \tag{IR}
\]

\(^{11}\)For a proof of the sufficiency of (1.12) and (1.13) see appendix A3.1 in Laffont and Tirole (1993). The proof of sufficiency requires some additional, but reasonable, assumptions regarding the cost function (1.1).
To solve program (3.13) it is best to maximise the objective function subject to the (IR) and (IC) constraint only and then examine whether the solution found satisfies condition (c3) and (c4). If it does, then the solution is feasible and the optimal contract has been found.\(^{12}\)

Note that we have changed the control variables from \(q, C, t\) to \(q, e, U\). Equation (1.1) shows that once \(q\) and \(e\) are determined so is \(C\), therefore it does not matter which control variable is used. Also, by equation (3.11) once effort and utility are determined so is the transfer from the regulator.

Taking \(e\) and \(q\) as control variables and \(U\) as a state variable, the first order conditions for program (3.13) are\(^ {13}\),

\[
\frac{dC}{d\theta} \geq 0 \\
\frac{dq_i}{d\theta} \leq 0 \quad \forall i.
\]  

\[(c3)\]

\[(c4)\]

Comparing these first order conditions with the case of symmetric information we see that the last term in each equation is an adjustment to the first best (symmetric information) conditions. These adjustment are due to the desire of the regulator to limit the informational rents accruing to firms. Rearranging equation (3.7) we obtain,

\[
\lambda F(\theta) \left[ \psi''(e^*) \frac{\partial E}{\partial \theta} + \psi'(e^*) \frac{\partial^2 E}{\partial \theta \partial C} \right] = (1 + \lambda) \left( -\frac{\partial C}{\partial e} - \psi'(e^*) \right) f(\theta).
\]

The right hand side of this equation is the social benefit of increasing effort (of a firm of type \(\theta\)) by one unit. It is valued at \((1 + \lambda)\) since cost

\(^{12}\)If it does not meet the conditions, the solution must be modified by "bunching" some types (see Guesnerie and Laffont (1984)).

\(^{13}\)See appendix A3.1 in Laffont and Tirole (1993).
reductions reduce the necessary public transfers given to the firm to cover its costs. Also, it is multiplied by the density associated with type $\theta$ firms. The left hand side is the extra rents that will have to be given to all firms that are more efficient than $\theta$. If the effort exerted by a type $\theta$ firm is increased, then more efficient firms will earn higher utility by pretending to be less efficient. Therefore, to maintain incentive compatibility, the mechanism has to award the extra utility to efficient types. The cost of the extra transfer is valued at $\lambda$ and there are $F(\theta)$ firms that are more efficient than $\theta$.

The optimal mechanism then trades-off efficiency gains, by lowering effort to more inefficient firms, in order to extract informational rents from efficient types. The mechanism has the property that, for the most efficient type, there are no additional efficient firms whose rents must be limited. Therefore the level of effort for the most efficient type is the first best level.

The adjustment to the resource allocation conditions (3.6) has a similar interpretation. The regulator may want to limit the output produced by a firm of type $\theta$. This would be the case if production affects the amount of effort that a more efficient type saves by pretending to be less efficient.

For the remainder of this and the following chapters, we will assume that the cost function has the following form

$$C = C(\xi(\theta - e), q), \quad (1.17)$$

which amounts to assuming that effort and the efficiency parameter are weakly separable from output in the cost function. The attraction of this assumption is that it is sufficient for the incentive-pricing dichotomy to hold (Laffont and Tirole (1990))\textsuperscript{14}. Assumption (3.15) implies that,

$$\frac{\partial E}{\partial \theta} = 1$$

\textsuperscript{14}Assumption (3.15) is stronger than necessary. A necessary and sufficient condition is that $C = C(\xi(\theta, e), q)$. 

29
and is independent of the cost level and the output vector. This assumption simplifies the analysis without losing any important insights. The optimal (second-best) mechanism would now be described by the following conditions,

\[
\frac{\partial V}{\partial q_i} = (1 + \lambda) \frac{\partial C}{\partial q_i} \quad \text{for } i = 1, \ldots, n \tag{1.18}
\]

\[
\psi'(e^*) = -\frac{\partial C}{\partial e} - \frac{\lambda}{(1 + \lambda)} \frac{F(\theta)}{f(\theta)} \psi''(e^*) \tag{1.19}
\]

\[
U(\theta) = \int_\theta^\bar{\theta} \psi(e^*) d\zeta. \tag{1.20}
\]

The incentive-pricing dichotomy is so named because the optimal allocation or pricing condition (1.18) is the same under symmetric and asymmetric information. The complications that arise due to asymmetric information are all dealt with in the "supply" side of the regulatory contract\(^\dagger\).

## 1.3.3 Remarks

The canonical model presented above does not consider many important issues dealing with regulation, or the possible variants that can be developed by relaxing some of the assumptions. Its purpose was to illustrate the basic problems facing regulators (asymmetry of information) and to characterise the optimal second-best contract. This serves as a useful framework to analyse different regulatory schemes and to organise the discussion regarding the different dimensions of the regulatory task.

Many extensions to the model can be found in Laffont and Tirole (1993) and the references and bibliographical notes found therein. We will not discuss those models here, but some particular issues will be touched upon further below.

---

\(^\dagger\)This not to say that output and prices would be the same under both informational scenarios. Since effort will be different under asymmetric information, so will marginal cost and therefore the optimal output will differ in both cases. Also, this result does not hold when costs are not observable as in the Baron and Meyerson (1982) model.
Of particular importance is the extension of the model to a multi-period setting and the introduction of dynamics. In a multi-period setting the inability of regulators to commit to a particular regulatory contract has dire consequences for optimal regulation. The basic problem is that optimal truth revealing incentive contract derived above is no longer incentive compatible. If the firm were to reveal its true efficiency type, \( \theta \), in the first period, the regulator could use this information in a subsequent period to design a contract that leaves no rent to the firm. Unless the regulator can commit not to change the contract in the future, the firm will be wary of revealing its true type.

1.3.4 Three essays in applied regulation

All of the essays in this thesis deal with some aspect of the regulatory model presented above and with equations (1.18), (3.16) and (1.20). The first essay deals with resource allocation issues. Equation (1.18) shows that demand issues are central to the regulatory task. The second and third essay are concerned with the supply side of regulation. In particular, the design of an optimal contract and the rent extraction aspects of regulation. Each of these topics and the relation with the ensuing essays will be further elaborated below.

1.4 The resource allocation side of regulation

Equation (1.18) determines the optimal pricing of output in a regulated industry. The issue of pricing has a long history in the regulatory literature. Due to the natural monopoly characteristics of these industries, competition could not be relied upon to establish social optimal prices in an unregulated context. Furthermore, historically many of the utility industries have been publicly owned (at least in the post-war years in the United Kingdom) and guidance was needed to determine how prices should be set.
Much intellectual effort was expended after the Second World War on determining optimal pricing and investment policies in public enterprises. Modern project appraisal techniques, peak-load pricing theory, and the concept of pricing at long-run marginal cost were developed at the time. Those theoretical advances had a direct influence on policy prescriptions—if perhaps not actual practice—in the United Kingdom, as the 1961 and 1967 White Papers on the nationalised industries show.

Most of this literature is still relevant to the New Theory of Regulation. In this section we will review that part of the pricing literature dealing with tariff structures. We will not cover dynamic or investment issues or peak-load pricing problems. Some extensions to the theory have also been developed to analyse issues raised by recent regulatory reforms. We will also briefly examine those results.

Before continuing it is necessary to state the conditions that make the pricing structures analysed below feasible. In most cases, optimal tariffs will involve positive price-cost margins and a declining average price for the regulated goods. The regulated firm must then have some market power if positive price-cost margins are not to be eroded by competition. The assumption of market power is usually the reason why these companies are regulated in the first place, so this assumption is imbedded in the original regulatory model. However, having market power, as in the case of a monopoly provider, is not enough to guarantee the feasibility of tariff structures that imply some kind of quantity discount. In this case, there must also be some obstacle to arbitrage between consumers. Otherwise, there would be a profit opportunity for one customer to buy large amounts at a discount and resell to smaller customers. Arbitrage possibilities between customers is prevented by the difficulty of storing some goods, such as the energy products of utilities such

\[ \text{See Brown and Sibley (1986) and Mitchell and Vogelsang (1991) for unified treatments of public utility pricing theory. Early contributions to the subject include Hotelling (1938), Coase (1946) and Boiteux (1956).} \]
as electricity or gas. In other cases, such as telecommunications, the idiosyncratic nature of demand may make it difficult to establish trade between consumers. My telephone calls are not generally at the same of time, day, duration and distance as my neighbour's calls. In the case of telecommunication, however, technological change and deregulation has made arbitrage a real possibility. An example is the leasing of trunk lines by intermediate providers. The increasing arbitrage possibilities in telecommunications would indicate that tariffs will increasingly be determined by competitive pressures. Later we will come back to this issue.

1.4.1 Linear prices with homogenous consumers

Now we can go back and examine what \( V(q) \) is. The simplest case to analyse is to assume that all consumers have the same preferences and they have an additively separable utility function for the vector of products \( q \) and other expenditure. We can therefore speak of a representative agent. In addition, we assume that each output is sold at a constant price \( p_i \) per unit. The social value of production would be

\[
V(q) = S(q) - p(q)q + (1 + \lambda)p(q)q \\
= S(q) + \lambda p(q)q
\]

where \( S(q) \) is gross consumer surplus from output \( q \), and \( p(q) \) is a vector of inverse demand curves. Revenue has a net social value since it reduces the amount of distortionary taxation that needs to be raised. We will make the simplifying assumption that there are no cross price effects in the demand curve of each good. Therefore, \( \frac{\partial S(q)}{\partial q_i} = p_i \) and \( \frac{\partial p(q)}{\partial q_i} = \frac{\partial p_i}{\partial q_i} \). In this case, condition (1.18) would be,

\[
\frac{p_i - \frac{\partial C}{\partial q_i}}{p_i} = \frac{\lambda}{(1 + \lambda)} \frac{1}{|\eta_i|}
\]

(1.21)
where $\eta_i$ is the demand elasticity of good $i$. The "Ramsey type" formula (1.21) states that price-cost mark-ups should be inversely proportional to demand elasticities\textsuperscript{17}.

In the present interpretation, price above marginal costs are due to the convenience of raising funds in this regulated industry rather than through distortionary taxation. In particular, $\lambda$, the cost of public funds, is exogenous to this partial equilibrium analysis. This case should not be confused with the case where a regulated firm is not allowed to receive direct transfers from the regulatory authority. In this later case, when there are increasing returns to scale the firm cannot cover all costs through the revenue received from marginal cost pricing. Therefore, to break even, the firm must increase prices above marginal costs. The optimal (welfare maximising) way to do this is by setting price-cost margins according to an equation identical to (1.21) except that, in this case, $\lambda$ would be a Lagrangian multiplier associated with the firm's budget constraint and would therefore be endogenous to the analysis.

The important point to note in this section is that it is socially optimal to set different price-cost margins for the different goods. If the goods are defined as outputs for, say, different regional markets or different customer classes, optimal linear pricing would indicate that some price discrimination among regions or customer might be desirable.

### 1.4.2 Two-part tariffs with homogenous consumers

Condition (1.21) is an optimal trade-off between the welfare reducing distortions caused by prices being above marginal costs (or output lower than necessary) and the welfare enhancing benefits of lower distortionary taxation. See Brown and Sibley (1986). Another interesting case to analyse is that of a private monopoly trying to maximise profits. In that case, the optimal pricing strategy is given by equation (1.21) but with $\lambda = \infty$.

\textsuperscript{17}If there are cross price effects in demand then an equation analogous to (1.21) obtains but the elasticity is replaced by a “super” elasticity that takes into account cross price effects. See Brown and Sibley (1986). Another interesting case to analyse is that of a private monopoly trying to maximise profits. In that case, the optimal pricing strategy is given by equation (1.21) but with $\lambda = \infty$. 

34
in other parts of the economy. However, welfare can be improved further if we relax the linear price assumption and allow for a two-part tariff structure. In this case the customer pays a fixed charge, \( a_i \), that is independent of the units consumed, and a variable price, \( p_i \), for each unit consumed. Formally, a two-part tariff is an outlay schedule

\[
P_i(q_i) = a_i + p_i q_i.
\]

The fixed charge can be used to raise revenue while allowing the unit price to be set equal to marginal cost, therefore reducing the distortions caused by price-cost markups.

With homogenous consumers and no income effects, the fixed charge acts as a non-distortionary lump-sum tax. Therefore, it is optimal to raise as much revenue as possible using the fixed charge. The social value of production is now,

\[
V(q) = S(q) - \left( \sum_{i=1}^{n} a_i + p(q)q \right) + (1 + \lambda) \left( \sum_{i=1}^{n} a_i + p(q)q \right)
\]

\[
= S(q) + \lambda \left( \sum_{i=1}^{n} a_i + p(q)q \right).
\]  

(1.22)

Net consumer surplus is \( S(q) - \sum_{i=1}^{n} a_i - p(q)q \). The maximum amount of revenue that can be raised with the fixed charges, consistent with consumers still purchasing some amount of the goods would be,

\[
\sum_{i=1}^{n} a_i = S(q) - p(q)q.
\]

Substituting this expression into (1.22), the social value of production would then be,

\[
V(q) = (1 + \lambda) S(q).
\]
It is easily checked that, in this case, the optimal pricing condition (1.18) requires marginal cost pricing,

\[ p_i = \frac{\partial C}{\partial q_i} \quad \text{(1.23)} \]

It is interesting to note that the above prices are exactly equal to what a perfectly discriminating monopoly would set (Oi (1971)). In both cases the firm extracts all of consumer surplus.

### 1.4.3 Two-part tariffs with heterogeneous consumers

The assumption of homogenous consumers is obviously a fiction. Replacing this assumption with that of heterogeneous consumers leads to different and interesting results. We must first clarify what is meant by heterogeneous consumers. One interpretation is to group consumers according to an observable characteristic. For example, students, non-students and pensioners. The demand behaviour of each customer group may be different and the optimal pricing question is: what would be the optimal charge to each group? This is a problem of third-degree price discrimination\(^\text{18}\). It turns out that the results of the last two sections apply to this case, provided that the subindex \(i\) is interpreted as indexing customer groups instead of products. For example, optimal linear prices would involve high price-cost margins for groups with low demand elasticity and vice-versa.

In contrast to observable customer groups, the focus below is on unobservable heterogeneity. Preferences, and thus demand behaviour, varies across the population. The firm, however, cannot distinguish individuals by their observable characteristics. It does know, however, how preferences are distributed in the population as a whole. This problem is more akin to the second-degree pricing literature.

With heterogeneous consumers it may no longer be optimal to set the variable price equal to marginal cost. This holds for a welfare maximising

\(^{18}\text{See Varian (1989).}\)
regulator (Feldstein (1972); Ng and Weisser (1974); Auberbach and Pellechio (1978); Schmalensee (1981)) as well as for a profit maximising monopolist (Oi (1971); Schmalensee (1981)).

There are two sets of reasons why cost reflective prices may not be optimal. The first is distributional. The fixed charge acts like a regressive lump sum tax on consumers. If poorer consumers have a higher weight in the social welfare function than richer consumers, then it would be welfare enhancing to trade-off some of the efficiency gains of a high fixed charge for the distributional benefits of lowering this charge.

Feldstein (1972) analysed the above situation and obtains the result that price-cost margins should be inversely proportional to the absolute value of the demand elasticity (as in equation (1.21)) and negatively related to the covariance between consumption of the output and the social marginal utility of income across households. If low income households consume less than average quantities of the output, then this covariance will be negative and, consequently, more revenue should be raised by increasing the price-cost margin rather than by the fixed charge. Auerbach and Pellechio (1978) extend Feldstein's analysis to the case where market penetration is endogenous. That is, for a given fixed and variable price, not all consumers find it in their interest to participate in the market. Feldstein's result on the optimal price-cost margin is then adjusted for this market penetration effect.

The second reason for not setting the variable price equal to marginal cost does not rely on distributive arguments. When market participation is endogenous, lowering the fixed charge increases the consumer surplus for infra-marginal consumers. This positive welfare effect may compensate the negative impacts of having to increase the price-cost margin.

To analyse the optimal two-part tariff, let us assume that consumer heterogeneity can be represented by a one-dimensional parameter $\beta$, with a continuous distribution function $H(\beta)$ with support $\beta \in [\underline{\beta}, \overline{\beta}]$. Utility for a
consumer type $\beta$ is given by

$$V(\beta) = U(\beta, q) - a - pq$$  \hfill (1.24)$$

where we have assumed for simplicity that there is only one good. For a given marginal price and fixed charge, a consumer of type $\beta$ will choose a consumption level $q^*(\beta)$ that maximises (1.24). Indirect utility is then,

$$V(\beta, p, a) = U(\beta, q^*(\beta)) - a - pq^*(\beta).$$  \hfill (1.25)$$

There is a consumer type $\widetilde{\beta}$, who is indifferent from purchasing positive amounts of the good and not purchasing at all. This type is found by solving $V(\widetilde{\beta}, p, a) = 0$. Given certain regularity conditions on the way $\beta$ affects the marginal utility of consumption, the marginal consumer will be a type $\widetilde{\beta}$ such that all types $\beta < \widetilde{\beta}$ do not purchase the good while all $\beta > \widetilde{\beta}$ consume positive amounts of the product. $q^*(\widetilde{\beta})$ is the consumption of level of the marginal consumer.

Assuming away distributional issues, the optimal pricing conditions are

$$\frac{p - \frac{\partial C}{\partial q}}{p} = \frac{\lambda}{(1 + \lambda) |\bar{\eta}|} \left( 1 - \frac{q^*(\widetilde{\beta})}{\bar{q}} \right)$$  \hfill (1.26)$$

and

$$\frac{a + \left( (p - \frac{\partial C}{\partial q}) q^*(\widetilde{\beta}) \right)}{a} = \frac{\lambda}{(1 + \lambda) |\bar{\eta}_N|}$$  \hfill (1.27)$$

where $\bar{\eta}$ is the price elasticity of demand holding $\widetilde{\beta}$ constant, and $\bar{\eta}_N$ is the elasticity of the number of consumers who purchase positive amounts of the good with respect to the fixed charge.

The left hand side of equation (1.27) is also a price cost mark-up. The cost of supplying an extra consumer is minus the profits that will be made
from the purchases of the new consumer. Therefore, both pricing conditions
can be regarded as Ramsey pricing formulas19.

As in the case of linear prices, equations (1.26) and (1.27) show that
it may be socially optimal to set price-cost margins differently for different
goods. If the goods are defined as outputs for, say, different regional markets
or different customer classes, optimal linear pricing would indicate that some
price discrimination among regions or customer might be desirable. However,
in the case of two-part tariffs, price margins not only differ across goods but
also between the fixed charge and the variable price. That is, it may not
be socially optimal to make tariffs cost reflective. At the optimum, the
fixed charge may not necessarily reflect the marginal customer costs and the
variable price the marginal cost of output.

Equation (1.26) also shows the importance of heterogeneity for the design
of two-part tariffs. If the demand of marginal consumers is equal to aver­
age consumption then there is no benefit from having a price different from
marginal cost.

1.4.4 Non-linear pricing

Compared with linear prices, optimal two-part tariffs cannot decrease the
level of welfare. This is because linear prices are a special case of two-part
tariffs. However, two-part tariffs are a special case of more general tariffs.
Therefore, any welfare level achieved by two-part tariffs can also be achieved
by more general multi-part tariffs and, usually, these last tariffs will be able
to do better20.

The highest level of welfare will be achieved by fully non-linear prices,

19 Sometimes there are costs that are related to supplying a consumer independent of
the amount purchased. An example is metering costs in the supply of gas and electricity.
If \( \frac{\partial C}{\partial q} \) is the marginal cost of supplying an additional consumer, then the numerator in the
left hand side of equation (1.27) is \( a - (\frac{\partial C}{\partial q} - (p - \frac{\partial C}{\partial q})q^*(\beta)) \).

20 Obviously this observations rests on the assumption that implementation costs do not
rise significantly as more complex tariffs are used.
whereby each consumer type faces a different price level. In many cases non-linear prices are equivalent to continuous quantity discounts for the purchase of the good. The quantity discount characteristics of two-part tariffs can be seen from the fact that the average price paid for the goods decreases as the amount purchased increases.

The optimal non-linear tariff, \( P(q) \), in the case of one good and where consumers differ in just one dimension, is given by a very similar formula to (1.21)\(^{21}\). To show this we will follow the formulation proposed by Wilson (1993). Wilson defines a demand profile, \( N(p(q), q) \), as the number of customers that purchase the unit \( q \) of the good at the marginal price \( p(q) = \frac{\partial F(q)}{\partial q} \) for that unit. Using the notation of the last section,

\[
V(\beta) = U(\beta, q) - P(q).
\]

Then, assuming the Spence-Mirrlees condition holds, namely \( \frac{\partial^2 U}{\partial q \partial \beta} > 0 \), the demand profile can be defined as

\[
N(p(q), q) = 1 - H(\beta)
\]

where \( \beta \) is given by

\[
\beta = \beta(p(q), q) = \frac{\partial U}{\partial q}(\beta, q) - p(q) = 0.
\]

The advantage of using the demand profile is that the optimal non-linear pricing schedule is given by the simple condition

\[
\frac{p(q) - \frac{\partial C}{\partial q}}{p(q)} = \frac{\lambda}{1 + \lambda |\eta_N|}
\]

where \( \eta_N = \frac{\partial N(p(q), q)}{\partial p(q)} \frac{p(q)}{N(p(q), q)} \) is the elasticity of the number of consumers that purchase the \( q \)th unit when the marginal price for the \( q \)th unit changes\(^{22}\).

Equation (2.1) has the same “Ramsey” property as the other optimal pricing rules discussed above. One way to think about condition (2.1) is to

---

\(^{21}\)See Armstrong (1996) and Wilson (1993) for the more demanding case of multiple products and multi-dimensional consumer types.

\(^{22}\)For this optimal tariff to be feasible it must result in a non-decreasing demand for higher \( \beta \) types. That is \( q^*(\beta) \) must be non-decreasing in \( \beta \).
consider each unit of \( q \) as a different good. Then the price cost margin should be inversely proportional to the demand elasticity of each good, analogous to result (1.21) above.

### 1.4.5 Pricing and regulation schemes

All the above sections assume that the regulator can control prices, either directly in the case of a public firm or indirectly through an optimal truth telling mechanism. However, many recent regulatory proposals give private regulated firms flexibility to set prices within certain limits. Price cap regulation is a case in point\(^{23}\). The regulator sets the maximum average price that the firm can charge for a basket of goods. For example, when all the products of a firm are measured in the same units, say kilowatt/hours, the cap may take the form of a limit on the average revenues a firm can earn. Provided the firm meets the average revenue constraint, it is free to set prices for individual goods within the basket as it finds convenient.

An interesting question then arises: what are the characteristics of the pricing structure that a profit maximising firm would set and what are its welfare properties? Armstrong and Vickers (1991), Sappington and Sibley (1992) and Armstrong, Cowan and Vickers (1995) analyse this issue.

Armstrong, et al. (1995) compare three possible tariff structures under average revenue regulation. The first is to force the firm to charge a uniform tariff, the second is to allow the firm to charge a non-linear tariff (provided that the average revenue constraint is met), and the third is to allow the firm to offer a non-linear tariff but also giving customers the possibility to buy at the uniform price.

The third case of optional tariffs is pareto improving over uniform prices. The welfare comparison between uniform prices and non-linear tariffs (without the uniform price option) is ambiguous. However, unless the price cap is

\(^{23}\)Price cap regulation, along with other popular regulatory schemes, are discussed in the next section.
set significantly above marginal cost, uniform pricing is to be preferred. On average, consumers are better off with a uniform tariff than with a non-linear tariff. This would indicate that regulators need to be careful when allowing a price-cap regulated firm to price discriminate.

This last result is relevant to issues discussed further below. The profit-maximising non-linear tariff, subject to an average revenue constraint, involves marginal prices below marginal cost for the high \( \beta \) types. The resulting increase in demand from these types eases the revenue constraint and allows the firm to increase revenues from other types.

### 1.4.6 Empirical applications and the first essay

The pricing theory reviewed above requires knowledge of, at least, the demand elasticities for different goods and services. Therefore, demand studies that generate this information are an important prerequisite to apply the above pricing conditions.

In the first essay of this thesis a household level demand system is estimated for the United Kingdom. Special attention is placed on domestic energy goods, gas and electricity. The demand system is a Quadratic Almost Ideal Demand System (QUAIDS). This is a recent extension to the popular Almost Ideal Demand System (Deaton and Muellbauer (1980)) developed by Banks, Blundell and Lewbel (1997) that allows for quadratic Engel curves.

One of the novelties of the model is the use of desegregated price data for the energy goods. These goods are sold in the United Kingdom using two-part tariffs. Information was obtained on the fixed and variable charges for gas and electricity. Standard microeconomic theory tells us that the effects on consumer behaviour of a change in the variable price for a certain good will not be the same as the response to changes in the fixed charge. Therefore, to obtain consistent results, the model must distinguish between both prices. Two-part tariffs also raise interesting econometric issues which must be dealt
with in order to obtain consistent estimates.

The information provided by the estimated model could be used for many purposes. For example, the estimated elasticities could be used to design optimal tariff structures along the lines discussed in the previous sections. However, in the first essay the model is used to address a different but related issue.

One of the results of the optimal pricing literature is that some level of price discrimination may be justified on economic grounds. For example, price mark-ups may differ according to the demand elasticities for each product. Also, the balance between the fixed component of tariffs and the variable price should not necessarily reflect the relative value of customer related costs and marginal costs.

In practice, tariffs in many regulated industries present some kind of discrimination or cross subsidy between consumer groups or between different services. These tariff structures may reflect optimal pricing considerations or they may be the result of historical and political events. The important point, however, is that recent regulatory reforms have rendered many of these cross subsidies unsustainable. Therefore, in many industries there is currently a pressure to phase away cross subsidies and make prices more cost reflective.

Tariff rebalancing, however, may have undesirable welfare consequences. Many regulated industries such as gas and electricity provide services that are considered to be basic needs. In telecommunications, universal access is considered a legitimate social policy objective. Regulators and policy makers are therefore concerned for the distributive and other effects that the elimination of cross subsidies may have. Few studies address these issues, a noticeable exception being Wolak (1996).

Wolak (1996) studies the telecommunications industry in the United States\textsuperscript{24}. He estimates two demand systems (a QUAIDS and a Translog)

\textsuperscript{24}See also Cain and MacDonald (1991) and Kaserman, Mayo and Flynn (1990).
to analyse the impact that a rise in local rates, in tandem with a fall in long distance rates, would have on different groups of consumers. Until recently, long distance tariffs in the United States have been used to cross subsidise local calls. However justified on economic or distributive grounds, these cross subsidies are becoming difficult to sustain in the face of technological change and intense competition among long distance carriers. These developments have increased the pressure to rebalance long-distance and local tariffs to reflect their underlying cost.

Wolak's results indicate that there would be little loss in welfare for households that are adversely affected by tariff rebalancing and that disconnections would not increase significantly. In the aggregate, net welfare gain is positive implying that those that gain from the tariff change can more than compensate those who lose.

The demand system estimated for the United Kingdom is used in a similar fashion as Wolak (1996). Since April 1996, competition for the supply of domestic gas has been gradually introduced in the United Kingdom with electricity expected to follow in 1998. The introduction of competition in the domestic energy supply markets will entail a gradual elimination of existing cross subsidies. In particular, the fixed charge for gas and electricity is likely to rise while the variable price decreases. According to Hancock and Waddams-Price (1995), the potential welfare impact of tariff rebalancing has already affected policy decisions by the gas regulator. Measuring these welfare effects are important to inform future policy decisions and, possibly, to design specific programs to curb the negative social effects that the changes in the energy markets will have.

The results of the analysis of chapter 2 shows that if current cross subsidies in the domestic gas market are unbundled, some households would suffer welfare losses as their overall gas bill would increase. Furthermore, the latter group is composed of poorer and more vulnerable households. We also show that if competition reduces customer related costs by between 20% and
30%, and these costs reductions are translated into lower fixed charges, then almost all households stand to gain. However, if the potential cost savings that may be achieved in a competitive market are below this 20% to 30% range, then tariff rebalancing sparked by the introduction of competition does pose distributional conflicts.

The analysis of the welfare impacts of tariff changes is important in its own right. Results can be used to evaluate the need for special programs to alleviate negative distributional impacts and to help design these programs if required. But, welfare analysis may also have more direct policy implications for regulators. In some cases, the threat of adverse welfare consequences may be used by a monopoly incumbent as an argument against proposed regulatory reforms. Analysing the veracity of distributional and welfare claims—as is attempted in chapter 2—may, therefore, be a vital input to regulatory reform in utility industries.

1.5 The supply side of regulation

Pricing issues dominated much of the literature on utility regulation until two decades ago. Except for the Averech-Johnson effect (Averech and Johnson (1962)) incentive issues were not the focus of attention. In particular, the effects that ownership, or a particular regulatory scheme, might have on the incentives to reduce costs were not well developed themes in the regulatory literature. The idea of setting prices at an optimal margin—possibly zero—above marginal cost was well understood. However, nothing guaranteed that the marginal costs were the lowest possible.

This began to change as economist began realising that incentive issues were probably more important than pricing issues. As Newbery states:

---

25 The Averech-Johnson effect states that firms regulated with a rate of return scheme have an incentive to be more capital intensive than what efficiency considerations would dictate. This model assumes that the allowed rate of return is exogenously set higher than the cost of capital of the firm.

45
"...although the principles can guide pricing and investment decisions provided the nationalised industries are efficiently managed and minimise cost of producing the desired levels of output, they do nothing to ensure that the industries are managed efficiently. Moreover, it is not difficult to argue that it is far more important to ensure that costs are minimised than that prices are set at the correct level. Suppose the elasticity of demand facing a nationalised industry is unity; then, roughly speaking, the welfare cost of setting the price 20 per cent too high or too low is no greater than the welfare gain of cutting production costs by 2 per cent; whilst if price elasticities are lower than unity, pricing errors are less costly. It is doubtful whether the disagreements over pricing amounted to more than 20 per cent, and I suspect that very few economists would have claimed that improved management, monitoring, or competitive pressure could not have been able to reduce costs by at least 2 per cent."\(^{26}\)

The regulatory model presented above takes explicit account of incentive issues. It attempts to model the fact that to increase efficiency the firm must exert costly effort and must therefore be given the incentive to do so. Also, it takes account of the fact that the regulator has an informational disadvantage vis-à-vis a private firm. Besides deriving the optimal contract in such a situation, the above model clarifies the properties of other regulatory schemes that are used in practice. That is the subject of the next section.

1.5.1 Implementation and menus of linear contracts

Equation (3.16) and (1.20), together with (3.11), determine the level of effort, cost and transfer, all as a function of the firm's announcement of \( \theta \), of the optimal regulatory contract.

Under certain circumstances, the optimal contract can be implemented using a menu of linear contracts. The requirement is that the transfer of the optimal contract as a function of realised costs, \( t(C(\theta, e^*, q^*)) \), be convex. We will not present the necessary technical conditions for convexity to hold\(^{27}\).

The optimal contract can then be implemented by offering firms a suitably designed menu of linear contracts. Each contract of the menu has the form,

\[
t = a - bC
\]

where \( a \) and \( b \) (\( 0 \leq b \leq 1 \)) are parameters. These linear contracts are tangent to the optimal transfer function. The parameters \( a \) and \( b \) are chosen according to equations (3.16) and (1.20) so that it is optimal for a type \( \theta \) firm to choose the contract that was designed for that type.

The parameter \( b \) determines the "power" of the contract. That is, it determines the incentives that a firm has to reduce costs. If the firm reduces costs by one unit the transfer it receives increases by \( b \). Therefore, it will have the incentive to increase effort until

\[
-b\frac{\partial C}{\partial e} = \psi'(e).
\]

More effort will be exerted for higher levels of \( b \). In the limit, when \( b = 1 \), the firm has the incentive to exert the first-best (full information) level of effort. From condition (3.16) it can be seen that this is the level of effort required from a firm type \( \theta \). For all other types \( b \) will be strictly smaller than one. The least efficient firm will choose the lowest powered contract (with the lowest \( b \)).

\(^{27}\)Technical treatment for a particular family of multi-product cost functions can be found in proposition 3.2 and appendix A3.2 of Laffont and Tirole (1993). The procurement case and the single-product case are treated in chapter 1 and 2 of that book.
1.5.2 Price-cap regulation, rate of return regulation and profit sharing

Expressing the optimal contract as a menu of linear contracts permits a simple comparison with other regulatory schemes. We will discuss three schemes that are applied in practice and compare them to the optimal contract.

In the United States many public utilities are private. To avoid the abuse of monopoly power, these firms have traditionally been regulated with rate of return regulation\(^{28}\). Under this scheme the regulator sets prices so that revenues are sufficient to cover costs and leave a “fair” rate of return for the capital invested in the firm. Apart from the difficulties of measuring the capital stock and determining the “fair” rate of return of these assets, rate of return regulation has the inconvenience that firms have very little incentive to reduce costs. If the firm is inefficient it will still earn enough revenues to cover these costs and obtain an attractive return on assets. On the other hand, cost reductions achieved through (costly) effort are passed on to consumers in the form of lower prices.

In terms of the language of the last section, rate of return regulation is akin to a linear contract with \(b = 0\). Net transfers from the regulator are independent of realised costs so utility for the firm is

\[
U = t - \psi(e) = a - \psi(e)
\]

which is maximised when \(e = 0\) for all \(\theta\) types.

Arguably the case of public firms can also be modelled as rate of return regulation\(^{29}\). If the central government covers the cost of the firm, say by

\(^{28}\)Rate of return regulation is sometimes referred to as cost-plus regulation.
\(^{29}\)A more explicit model of the control of a public natural monopoly can be found in Rees (1989).
subsidising the firm if it makes negative profits, the managers will have very low incentives to reduce costs.

Due to the poor incentive properties of rate of return regulation, recent regulatory reforms have attempted to use more “powerful” regulatory mechanisms. In the United Kingdom this has taken the form of price-cap regulation\(^\text{30}\). The price cap consist of a limit to the average price or revenue that a firm can earn in a given period. Sometimes the cap is linked to the average revenue for a basket of goods, leaving the firm freedom to set the relative price for each individual product as it wishes. In the United Kingdom, the cap is set for a number of years and is expressed as an \( RPI - X \) formula. The average revenue for a firm can increase from one year to the next at a rate equal to the Retail Price Index (RPI) minus some factor \( X \). This last factor is an estimate of the real price reductions that can be achieved consistent with the firm being profitable\(^\text{31}\).

Once the price cap is determined, and ignoring possible strategic behaviour geared to influence the regulator at the next rate setting, the firm will have incentives to lower costs. Any efficiency gain that reduces costs will result in an increase in profits. During the period between rate reviews prices are not sensitive to costs and the firm becomes residual claimant to any cost reductions.

Price-cap regulation is similar to the regulator offering all firms a linear contract with \( b = 1 \). All firms, including high \( \theta \) types, have the incentive to increase effort to the first best level.

As can be seen, price-cap and rate of return regulation can be thought of as two polar cases of linear contracts, with \( b = 1 \) or \( b = 0 \). There are also

---

\(^{30}\)See Armstrong, Cowan and Vickers (1994) for an extensive review of the British experience with price-cap regulation. See also Crew and Kleindorfer (1996) for a comparative review of the experience with incentive regulation in the United States and the United Kingdom.

\(^{31}\)In practice, rate of return considerations enter the determination of the price-cap and especially the \( X \) factor. Therefore, in a dynamic context, the difference between price-cap and rate of return regulation is more of emphasis.
intermediate possibilities. For example, the regulator could offer the firm a cost reimbursement rule whereby the firm gets to keep only a fraction of cost reductions. On the other hand, if costs are higher than originally envisaged, the regulator reimburses part of the extra costs to the firm. Although cost sharing or profit sharing schemes in practice are usually more involved than this, for the present purpose we can think of these schemes as linear contract with \( 0 < b < 1 \).\(^{32}\)

All three types of contracts have been used in practice to regulate natural monopolies. Compared with the optimal contract, all three mechanisms are suboptimal due to the fact that they offer one contract to all types of firms. We have seen that the optimal regulatory scheme offers a menu of contracts such that firms self-select according to their efficiency level.

1.5.3 Applicability of the optimal contract

Any real world regulatory situation is inherently dynamic. In addition, most regulators are not able to commit to a regulatory contract for a long period of time, if at all. As mentioned above, dynamics, coupled with a lack of commitment capacity on the part of the regulator, creates tremendous difficulties for the design of an optimal mechanism. The truth revealing regulatory contract derived above is only valid in a static context.\(^{33}\)

Even in a situation where we assume away dynamic problems – by assuming, for example, that the regulator can commit not to use the information acquired in early periods to impose a first best contract in later periods – and renegotiation of the contract is not possible ex-post, the amount of

\(^{32}\)Cost can vary due to factors that, unlike effort, are outside of the firms control. Because cost (or profit sharing rules) reduce the variability of transfers they are important when analysing regulation under exogenous cost uncertainty (see Schmalensee (1989)).

\(^{33}\)The contract is also optimal in a repeated context if the regulator can commit. However, the contract is not renegotiation proof, as both the regulator and the firm could gain by rewriting the contract once information is revealed. See Laffont and Tirole (1993) for a renegotiation proof contract with commitment.
information needed to design an optimal contract seems vast\textsuperscript{34}.

The design of an optimal contract requires knowledge of the way that effort and efficiency enter the cost function. Effort is a particularly difficult variable to handle. Neither the level of this input nor its price, $\psi'(e)$, are observed, thus precluding traditional dual cost function estimation approaches. Moreover, the optimal contract depends on the perceived distribution of efficiency types, $F(\theta)$, which may be a subjective-and possibly wrong-belief on the part of the regulator.

However, the simple message from the theory-that a menu of suitably designed contracts may be a better way to regulate firms than a single contract-seems to be gaining some practical recognition. If one of the options offered in a menu is the past contract used to regulate the firm, then the individual rationality constraint will necessarily be met. If the other options are better, from a welfare viewpoint, than the past contract, then the menu can at worst achieve the same level of welfare as the past contract and it will usually do better.

According to Laffont and Tirole (1993), offering agents a menu of options is common in managerial compensation schemes. The difficulties faced by shareholders in controlling managers is a principal-agent problem analogous to the regulatory problem posed above.

Laffont and Tirole (1993) also discuss two experiments with menu contracts designed according to theory presented above. One was commissioned by the German Department of Defence and is an application of menu contracts in the procurement context (Reichelstein (1991)). The other is the design of a menu of contracts to regulate the Palisades generating plant in Michigan.

Another example, taken from Sappington (1996), illustrates well the ex-

\textsuperscript{34}Another way to assume away dynamic problems is used in the second essay below. It consists of assuming that each period the firm draws a new efficiency parameter from the same distribution and that this draw is independent of the firm's past efficiency. Basically, the regulatory problem is then just a series of static problems.
The explicit use of menus in a regulatory context. It refers to the regulation, by the Federal Communications Commission (FEC) of the United States, of local exchange carriers (LEC). A local exchange carrier is the company that provides local telephone services in a geographical area. Companies that provide long distance telephone services must use the LEC in order for customers to make or receive long distance calls. This gives rise to an access charge by the LEC for the use of the local lines. In 1991, the FEC introduced a regulatory scheme for the access charge based on firms selecting among two options. If a LEC agrees to reduce their real average access charge by 3.3% annually, then they are allowed to keep 50% of the earnings that produce a rate of return on capital between 12.25% and 16.25%. All earnings that produce a rate of return on capital above 16.25% are returned to consumers, while all earnings below 12.25% are kept by the firm. However, if the LEC agrees to reduce the access charge by 4.3% annually, then the 50% profit sharing range is increased to between 13.25% and 17.25%. With these options, higher profits can be earned by firms willing to reduce prices by a larger amount. This will be easier to achieve, and therefore more profitable, for the more efficient firms. Therefore, the options serve as a screening mechanism whereby more efficient firms choose the option that requires them to make more cost reducing effort and offer lower prices. To make this work, the second option must offer higher profits.

The above examples show that menu contracts based on the insights of the new theory of regulation are slowly being applied. However, in order to design the optimal contract much information is required.

Even when optimal contracts are not feasible, the new theory of regulation offers a useful setup to compare the welfare properties of simpler non-optimal regulatory schemes. For example, regulators might be interested in knowing which of two regulatory schemes (say cost-plus regulation or a price cap) should be preferred in a given situation. Much the same information as that needed to design the optimal contract is required to make this welfare
comparison. Several studies have proposed methods to generate this information. These studies, as well as other empirical studies of regulation, will be reviewed below.

1.5.4 Review of the empirical literature on regulation

Different empirical approaches have been used to study issues dealing with regulation. This methodological heterogeneity can be explained by the different objectives of each study and the availability of data.

Reduced form econometric models

When a researcher does not model the complete regulatory environment, but is interested instead in determining the sign and size of the relation between certain observed variables, the model can be classified as a reduced form model. An advantage of reduced form models is that by only examining the correlation between variables, specification errors introduced by structural modelling are avoided.

Reduced form model can be very useful to examine testable hypothesis of a theory. For example, regulatory theory predicts that firms regulated by a price-cap mechanism would exert more effort, and therefore produce at a lower cost, than firms regulated by a cost-plus arrangement. Given a suitable data set, this proposition could be easily tested using regression analysis.

An example of a reduced form econometric study is Mathios and Rogers (1989). They examine tariff data on intra-state long distance calls in the United States. Traditional price regulation for this industry was based on rate-of return regulation. However, after 1983 some states introduced some form of price-cap regulation. This gave firms greater price flexibility (below the cap) and a greater incentive to reduce costs and innovate. Mathios and Rogers (1989) estimate a regression with tariffs as the dependent variable and demand, cost, political and regulatory variables as independent variables. The principal variable of interest was a dummy which took a value
of one if the state allowed some form of pricing flexibility in the form of a cap and zero otherwise. The results show that those states where price caps were introduced had lower tariffs than states that maintained rate of return regulation. However, it is difficult to interpret their results. Although theory would predict that price cap regulation would give firms greater incentives to reduce costs and to innovate, there is no convincing reason why the firms should subsequently reduce prices (in most cases below the maximum level allowed by the cap). It may be possible that the study is plagued by an endogeneity problem. Regulators may be more willing to introduce pricing flexibility if competition has already taken hold in the market and the incumbent needs more flexibility to compete. While in those states where a dominant firm has overwhelming market power, the regulator may be unwilling to abandon the rent extracting advantages of rate of return regulation. Although only a hypothesis, the above argument would imply that the price cap dummy in the Mathios and Rogers (1989) study is proxying the development of competition in each state and does not say much about the impact of different regulatory schemes.

Unlike Mathios and Rogers (1989), Dalen and Gómez-Lobo (1995) compare the impact of different regulatory schemes on costs rather than prices. They estimate a translog dual cost function (along with input share equations) for the Norwegian bus industry. Among the regressors is a dummy variable representing the type of regulation faced by each bus company. Their results show that firms that face a type of regulation, that gives, in theory, more incentives for cost reduction, do in fact have lower costs. Gagnepain (1996) analyses the French urban bus industry in a similar manner to Dalen and Gómez-Lobo (1995). He also shows that costs are lower when firms are regulated with more powerful regulatory contracts.

The reduced form approach is useful to analyse many hypothesis regard-

---

35 Details of the different regulatory schemes used in the Norwegian bus industry may be found in chapter 3.
ing regulation. However, it is not suitable for generating the information needed to design optimal contracts, to compare the welfare property of different regulatory schemes, nor in most cases to predict the effects of changing the underlying structural environment. In reduced form models, all interaction are subsumed in a “black box” regression that makes it difficult to identify crucial information such as, for example, the distribution of the efficiency parameter $\theta$.

**Parametrised, calibrated and simulated models**

In order to answer some of the questions that reduce form approaches cannot tackle, some authors have resorted to parametrised models. In these studies no use is made of econometrics\(^{36}\). Rather, a structural model and a grid of values for the different parameters of the model is specified. The model is then simulated under different regulatory schemes and for each set of parameter values. Welfare, effort, costs and other variables can then be analysed as parameters change.

Schmalensee (1989) takes this approach. He specifies a simple model for a single product risk neutral monopolist. Marginal costs are a function of pre-reform costs (historical costs), cost reducing effort and a random ex-post shock. The random shock is introduced to reflect cost uncertainty due to factors that are outside the firms and the regulators control. Asymmetric information between the firm and the regulator takes the form of a “type” parameter that affects the cost of effort of the firm.

Schmalensee (1989) limits his analysis to linear pricing regimes, where prices are a linear function of observed costs. Therefore, he does not consider the optimal menu discussed above, but only to single contracts with a particular “power” parameter.

The reason to limit the analysis to only linear regimes is twofold. On the

\(^{36}\)Wunsch (1994) and Gasmi, Laffont and Sharkey (1996), however, combine econometric estimates with calibrated functions.
one hand, Schmalensee argues that these regimes are, if not optimal, "good" in the sense that they are robust to errors in beliefs, they are easily understood and put in practice. He notes that most incentive schemes observed in practice are linear. The second reason has to do with technical matters. With ex-post cost uncertainty the optimal truth revealing contract is more difficult to derive than the model presented above.

Schmalensee (1989) compares the welfare properties of price-cap, cost-plus and the optimal linear contract using a grid of 625 points for different values of the demand elasticity, the distribution of expected cost savings, and the distribution of the exogenous cost shock. He finds that when ex-post cost uncertainty is present, price-cap regulation is almost never optimal. Price-cap regulation gives greater incentives for cost reduction but if there is uncertainty the cap must be set at a high price for the individual rationality constraint to be met (non-negative profits). On the other hand, if there is no ex-post cost shock then price-caps are generally optimal.

Gasmi, Ivaldi and Laffont (1994) take the same grid of parameters as in Schmalensee (1989) and compare the "best" linear contract to a price-cap regime that allows for downward price flexibility, a regime that combines a price-cap with a profit sharing rule, and the optimal asymmetric information contract. In their model they assume there is no cost uncertainty. They find that a price-cap regime leaves substantial rents to firms, but when this particular regime is appended with a profit sharing rule, the expected welfare is very close to the optimal non-linear contract.

Although the studies mentioned above are very useful to obtain qualitative results regarding the properties of different regulatory contracts, they are more limited as a tool for designing contracts in a particular real world situation. The studies have used very simple stylised cost and demand functions. In addition, there remains the empirical problem of determining which set of parameter values from the grid is applicable to a given situation.

A more promising approach is taken by Gasmi, et al (1996). They use an
engineering process model developed by Gabel and Kennet (1991) to simulate the costs of a local telephone exchange company for different efficiency, effort and output levels\(^{37}\). Technological efficiency, \(\theta\), is modelled as a multiplicative parameter on the cost of capital while effort is modelled as a multiplicative parameter on the price of labour. They assume that \(F(\theta)\) is a uniform distribution in the range \([0.5, 1.5]\) with 1 being the average cost of capital.

Gasmi, et al (1996) fit a translog cost function to the data obtained from the engineering model, and use the estimated results as the cost function \(C(\theta, e, q)\). The cost of effort function is assumed to have a simple quadratic parametric form. This function is then calibrated using anecdotal evidence regarding labour reductions after the privatisation of some telephone utilities around the world\(^{38}\). An isoelastic demand function is also calibrated using estimated elasticities from other studies. As the cost of public funds, they assume a value of 0.3.

With these calibrated functions and parameters, they compare the properties of five different regulatory schemes: optimal regulation under cost observability (Laffont and Tirole (1986)), optimal regulation when costs are not ex-post observable (Baron and Meyerson (1982)), cost-plus regulation, pure price-cap regulation, and price-cap with a profit-sharing rule.

**Structural econometric models**

Parametrised models have the disadvantage that they do are not based on real world data, and it is not clear how those models can be applied to a particular situation. The method used by Gasmi, et al (1996), on the other hand, depends on the availability of a suitable engineering model to generate cost data. These approaches may be useful for many purposes

---

\(^{37}\)The engineering process model optimises the telephone network according to the size, output, input prices and other characteristics given by the researcher.

\(^{38}\)They cite evidence that British Telecom in the United Kingdom reduced its workforce by 40% after privatisation.
and, in cases where there is sparse statistical data available, they are the only viable methods. However, some researchers have attempted another approach altogether, the use of structural econometric models.

Unlike reduce form models, structural models attempt to model all the interactions that generate the data in a particular situation. The empirical specification in a structural approach follows a relevant theoretical model closely.

In the regulatory context, Wolak (1994) seems to have been the first to apply a structural model to real world data. His study is interesting not only because he is able to recover from data the distribution of the efficiency parameter \( \theta \), but also because it highlights potential biases that may plague estimations that do not take into account the asymmetric information characteristics of the relationship between firms and the regulator.

Wolak (1994) uses the Baron and Meyerson (1982) approach to model the California water industry\(^{39}\). He specifies the cost function of the industry under two alternative informational assumptions. The first assumes that the regulator can observe the firms efficiency level and, therefore, the regulator imposes the first best (symmetric information) contract. The second assumes that the regulator has asymmetric information, and can only offer firms the optimal second-best incentive contract.

The two informational scenarios imply different reduced form cost functions. Wolak (1994) estimates both cost functions (together with the demand function to increase efficiency) using maximum likelihood techniques, and allowing for a non-parametric estimation of \( F(\theta) \). He then uses a non-nested specification test to analyse which informational scenario is more consistent with the data. The asymmetric information model is found to be a superior description of the data than the symmetric information model.

One important conclusion of Wolak’s study is that to estimate the cor-

\(^{39}\)Therefore there is no effort variable.
rect technological parameters in a regulated industry, the whole regulatory process should be explicitly modelled. For example, traditional cost function estimates will overstate the presence of scale economies when the regulator, faced with asymmetric information, offers the firm an optimal contract. This result, first shown with simulated data in Feinstein and Wolak (1991), can be easily explained. From (1.1), and remembering that in the Baron and Meyerson (1982) model there is no effort, the cost function to estimate will be,

\[ C = C(\theta, q). \]

The cost function is usually specified in such a way that the efficiency parameter affects costs (or a transformation of costs) in a linear manner\textsuperscript{40}. Therefore, the cost function to estimate will be,

\[ C = C(q) + \theta + \varepsilon = C(q) + \nu \]

where \( \varepsilon \) is a residual, and \( \nu = \theta + \varepsilon \). A truth revealing mechanisms is a set \( \{ q(\theta), C(\theta), t(\theta) \} \). Therefore, if the regulator is offering an optimal truth revealing contract to the firm, \( q \) will be a function of \( \theta \) and the cost function will be,

\[ C = C(q(\theta)) + \nu \]

whereby the correlation between \( q(\theta) \) and \( \nu \) will lead to a traditional endogeneity bias. The intuition for this result is quite simple. The regulator, after offering the truth revealing mechanism to the firms will know the efficiency type \( \theta \), and will command more efficient firms to produce more. Because they are more efficient, these firms will at the same time have lower costs. Therefore, low cost firms will be associated with high production levels. A

\textsuperscript{40}To simplify the exposition, in the following discussion we do not include input prices in the cost function.
traditional cost function estimation will attribute this to scale economies. However, in reality it is due to the endogeneity of output produced by the contract offered by the regulator.

The bias in scale economies discussed above is an example of the implications that the new theory of regulation may have for empirical analysis. Reduced form models may give seriously biased results and the use of the theory in a structural model may be a more reliable estimation method.

Thomas (1995) is another example of a structural econometric model applied to a regulatory situation. He models the regulation of water pollution abatement in France. In that country there exists a Pigouvian emissions tax for water pollution. For reasons outside the model, the Pigouvian tax is not set at its optimal level and the regulator must also rely on direct contracts with the firms concerning their investment in pollution abatement technology. Because of asymmetric information concerning the marginal abatement cost, the regulator offers an optimal truth revealing contract to firms.

The structural model is then estimated by a simulated method of moments technique under the assumption that θ (in this case a multiplicative parameter on the cost of abatement function) is log-normally distributed. The estimation method also takes into account the truncation in the distribution of θ that generate the data, since only the more efficient firms are regulated by contracts.

Thomas (1995) finds that the actual pollution tax is only 50% of the optimal Pigouvian level. In addition, he is able to determine the relative efficiencies in pollution abatement by industries, with the chemical industry being the most efficient.

Both Thomas (1995) and Wolak (1994) assume that, during the period when the data was generated, the regulator was offering an optimal contract under asymmetric information conditions. Although in the case of Wolak

\[41\text{Thomas (1995) shows that in the presence of the Pigouvian tax, it is optimal to regulate by contract only the more efficient firms in the industry.}\]
(1994) a formal test is presented against an alternative specification, the assumption of such a sophisticated regulator may be unrealistic in many situations.

Wunsch (1994) takes an approach that is closer to the second essay of this thesis. He uses data on 177 mass transit firms in Europe to design optimal contracts in the Laffont and Tirole (1986) setup. He does not assume that regulators are offering firms optimal contracts, but rather that regulation is low powered and therefore effort is low. The cost of effort function, \( \psi(e) \), is assumed to be quadratic and is calibrated comparing costs and profits of British transit firms (which are assumed to operate in a competitive environment) to those of continental firms\(^{42}\).

Wunsch proposes an interesting idea to calculate the distribution of efficiency \( F(\theta) \): "The main idea is to circumscribe the uncertainty about the unknown cost parameters by using regular inference techniques in econometrics. More precisely, the asymmetry of information between the regulator and the agent is assumed to be limited to the unexplained variance of a cost estimation based on a cross-section of 177 transit firms." Therefore, he estimates a cross-section cost function conditioned on the characteristics of each transit system. He obtains that the confidence interval around fitted values has a standard error of about 15% of costs. For \( F(\theta) \), Wunsch (1994) then assumes a normal distribution with this standard error.

The combined results from the econometric estimations and the calibration are then used to design optimal contracts and to analyse the gains from introducing this contract in a particular situation.

1.5.5 The Norwegian bus transport study

In the second essay of this thesis, data on the Norwegian bus industry is used to illustrate an approach that goes beyond the studies reviewed in the last

---

\(^{42}\)Gasmi, et al (1996) use this same idea to calibrate the cost of effort function.
sections.

The method follows the structural modelling approach to recover underlying parameter values. Unlike Wolak (1994), however, the study does not assume that the regulator is offering optimal contracts during the period when the data was generated. On the contrary, one of the objectives of the essay is to attempt to recover econometrically the information needed to design an optimal contract. The approach relies on the fact that Norwegian bus companies have been regulated by different schemes in recent years, and we use this regulatory heterogeneity to identify some of the parameters. The study also differs from Wunsch (1994) in that no calibration is used, and all the information is identified econometrically.

Another difference with Wolak (1994) is that the model assumes ex-post observability of costs by the regulator. This allows an effort variable to be specified in the cost function. One interesting finding is that we also obtain that conventional estimates of economies of scale will be biased, analogous to that first shown by Feinstein and Wolak (1991). However, in this case the bias is not a consequence of the optimal contract offered by the regulator, but rather due to ignoring the effects of effort in the cost function. Together with Wolak (1994), this result casts doubts as to the reliability of traditionally estimated scale economies in regulated industries.

The reason why ignoring effort can create a bias in econometric estimates is the following. If one is willing to assume (as is done in chapter 3) that firms with higher costs will gain more from cost reducing effort, then any increase in output or input prices that led to higher costs would also increase the productivity of effort. Firms would then have an incentive to increase effort and thus reduce average costs. In the data one would observe high output firms operating at lower costs than would be the case if effort was constant (or zero) and this lower cost would be attributed to higher economies of scale. Similar effects bias the coefficients on input prices.

As far as we are aware, the essay in chapter 2 is the only study so far that
allows optimal contracts to be designed relying exclusively on econometric estimates.

The model is used to make predictions of the effects of changes in the regulatory environment, an advantage of structural modelling. In particular, we make predictions about the savings in transfers that the Norwegian authorities can expect if tendering of bus routes is introduced.

The modelling exercise also clarifies under what conditions and what assumptions (especially functional assumptions) are needed in order to identify all the relevant parameters. Therefore, it contributes knowledge about the limitation encountered by structural empirical models of regulatory theory.

1.6 Rent extraction and auctions in regulation

The optimal second-best contract given by (1.18), (3.16) and (1.20) leaves positive rents to all but the least efficient firm. The contract must give up rents in order to induce more efficient firms to reveal their efficiency type. It may seem that these socially costly informational rents are unavoidable if the regulator has asymmetric information. However, when there is a multitude of firms that could potentially be the regulated monopoly, there are other regulatory mechanisms available that leave less rents than the optimal contract analysed above. In a natural monopoly situation it is not possible to have competitive provision of the goods, however, it may be possible to have ex-ante competition for the right to be the monopoly producer. This competition takes the form of an auction or tendering process.\footnote{When there are multiple firms producing in regional markets it may also be possible to use yardstick competition schemes. See Shleifer (1985).}

The idea of auctioning monopoly rights goes back to at least Demsetz (1968). Real world examples of auctions as a regulating mechanism abound. For example, most governments use a procurement process to buy the goods...
and services that it requires. The tendering of contracts for the construction of infrastructure, such as highways, is a more specific example. Some countries have even tendered the right to manage key public utility industries. For example, as an alternative to full-fledged privatisation, Argentina tendered concessions that last 20 years to run its electricity firms.

In this section we will briefly review the rational for introducing auctions in regulation, review the theoretical literature on auctions of incentive contracts, and discuss the growing literature on the econometrics of auctions. Before proceeding, however, it must be mentioned that auctions are not a panacea for all situations plagued with asymmetric information. In certain cases, the periodical auctioning of a regulatory contract may have negative consequences for investment in firm specific assets. A concessionaire may not be willing to invest in assets if he is not assured that he will be compensated for this investment in the event that the concession changes hand in the future. The valuation of assets at the time that a concession ends will be problematic when these assets are firm specific. The uncertainty faced by an investor may lead to under-investment or a bias against long-lived assets when periodical auctions of regulatory contracts is used. The alternative of auctioning the contract only once is not very attractive. As time passes, the concessionaire who won the original contract may not be the most efficient firm. Technological change may render the contract obsolete. In the end, the situation reverts to a game between an uninformed regulator and a natural monopoly as modelled above.

1.6.1 Auctioning incentive contracts

Laffont and Tirole (1987) derive the optimal auction for the right to be the monopoly producer in the asymmetric information regulation model pre-
sentiment above. Let's assume there are \( m \) bidders who independently draw their efficiency parameter \( \theta_i \) from the same distribution \( F(\theta) \). In the one period model, the auction that maximises welfare awards the monopoly franchise to the firm that announces the lowest efficiency parameter, \( \theta \). The winning firm is compelled to exert a level of effort given by (3.16), just like in the case of optimal regulation without an auction. The only difference is that the rents earned by the winning firm is given by

\[
U(\theta_1) = \int_{\theta_1}^{\theta_2} \psi'(e(\xi))d\xi
\]  

(1.30)

where \( \theta_1 \) is the announcement (bid) of the firm with the lowest efficiency parameter and \( \theta_2 \) is the second lowest efficiency type. The rents that the regulator must concede under an auction are smaller than in the optimal regulatory case. Examining equation (1.20) we see that the equivalent rents for a type \( \theta_1 \) firm under an optimal contract would be

\[
U(\theta_1) = \int_{\theta_1}^{\theta_2} \psi'(e(\xi))d\xi.
\]  

(1.31)

Thus the auction serves to reduce rents. In essence, the auction truncates the distribution of types from \([\theta, \bar{\theta}]\) to \([\theta, \theta_2]\). This truncation from above does not affect the optimal effort condition since the hazard rate, \( \frac{f(\theta)}{F(\theta)} \), is not affected by this truncation.

As the number of bidders increases, the first and second lowest valuation will converge to \( \theta \) thereby reducing informational rents to zero. This is the key attraction of auctions: as the number of bidders increase the regulator is able to overcome the problems posed by asymmetric information and the first best (symmetric) outcome is achievable. Regulators would presumably be interested in implementing auctions whenever possible. However, as discussed

---


46Notice that if \( \theta \in [\theta, \theta_2] \) the hazard rate would be \( \frac{f(\theta)}{F(\theta_2)} = \frac{f(\theta)}{F(\theta)} \) just as before.
above, for many sectors with long lived capital and sunk assets, auctions may not be feasible\textsuperscript{47}.

Laffont and Tirole (1987) propose a Vickrey type auction to implement the optimal auction. Firms declare an efficiency type $\theta$, the firm with the lowest declaration wins and is given a transfer according to (1.30). Firms would have a dominant strategy to bid their true type.

The optimal auction is not a typical auction observed in practice. Here the auctioneer would monitor ex-post costs and share part of the firms overrun costs\textsuperscript{48}. Most auctions observed in practice are fixed price auctions whereby a quantity of the good is offered in exchange for the highest bid price.

As confirmed by the results of chapter 3 on the Norwegian bus industry, the optimal auction will be very similar in welfare terms to a fixed price auction. The truncation of the distribution of types implied by the optimal auction means that the winning bidder will probably be quite efficient and will therefore be assigned a high effort level according to equation (3.16). In practice, this implies that the winning bidder will face a high powered incentive contract. Although only for the most efficient type would this contract be a fixed price contract, it will be very close to a fixed price contract for most of the winning firms.

Since the optimal auction and a fixed price auction will result in similar welfare levels, the added complexity of designing an optimal auction may not be justified. Therefore, although optimal auctions are discussed in the second essay, the focus of the following discussion, as well as the third essay, will be on simpler fixed price auction forms.

\textsuperscript{47}See Laffont and Tirole (1993) and Williamson (1976) for more on this topic.

\textsuperscript{48}However, auctions where the payment is linked to some ex-post variable are observed in practice. For example, the auction of publishing rights often include a royalty according to sales. These auctions could be interpreted as an attempt to apply the optimal incentive contract auction, although there are other complementary reasons why these royalty auctions might be convenient (see the discussion in McAfee and McMillan (1987)).
From a regulatory perspective, the study of auctions is important for the design of regulatory mechanisms. Auction theory provides some guidance on the reforms that may improve an auction mechanism under given circumstances. One example, is the determination of the optimal reservation price that a revenue maximising seller should set.

We will first review the econometric literature on auctions and we will then present an overview of the third and final essay. Because of space limitation, in what follows we will not cover the experimental auction literature nor the studies of optimal bidding from the research operation literature. For the first, see part IV of Smith (1991) while Laffont (1997) presents a brief survey of the second.

1.6.2 Non-structural empirical studies of auctions

There are several reasons for studying auctions empirically. One is the potential for testing certain propositions derived from auction theory. In contrast with other economic situations, auctions have the attraction that the “rules of the game” are known in detail and they correspond closely to the assumptions of the theoretical models. Therefore, auctions offer a promising context for testing game theoretic models and predictions (Bresnahan (1996); Laffont (1997)). As Laffont states “…despite the relatively small importance of auctions in economic activities, the methodological progresses which will be made in the future of the econometrics of auctions will be a useful indicator of what game theory can hope to bring to empirical industrial organisation”49.

One branch of empirical auctions literature attempts to test some implication of auction theory using reduced form models. These models usually specify a regression equation with the value of bids as the independent variable and the number of bidders, proxies for the value of the good sold, as well as other conditioning variables, as explanatory variables.

One popular topic using reduced form models has been to test for the existence of the winner's curse. This phenomenon may arise when bidders face uncertainty regarding the (common) value of the good for which they are bidding. If each bidder makes an unbiased estimate of the good in question, then the individual who has the most optimistic estimate of the good will, under a naive bidding strategy, present the highest tender. The winner will then pay a bid higher than the true (common) value of the good. This, in very simple terms, is called the winner's curse. To avoid it, bidders should optimally shed their bids to compensate for the fact that expected value of the good, conditional on winning, is lower than the unconditional expected value.

The winner's curse becomes more pronounced with more bidders since winning becomes an even stronger indication that the winner's estimate was biased. There is a conjecture, which has been shown to hold for specific cases, that as the number of bidders increase the optimal bidding strategy is to reduce an individual bid. Therefore, if one examines bids a number of (similar) auctions one would expect a negative relation, or an inverted U shape, between the value of individual bids and the number of bidders in each auction.

Studies that have examined this issue include Hendricks, Porter and Gertler (1986) and Gilley and Karels (1981). The latter correct for a possible selection bias when examining the relation between bids and the number of bidders. They use data on the bidding for offshore oil rights in the United States. They assume that there is a fixed cost to development of an oil

---

50See Rothkopf (1969) and Wilson (1969, 1977) for a theoretical analysis of the common value auction. See also the discussion in chapter 4. The other theoretical paradigm is the independent private value auction which will be discussed further below. Milgrom and Weber (1982) present the general case of affiliated values, which encompasses the common and private value paradigm as special cases.

51The expected winning bid, however, would still be positively related to the number of bidders.

52Earlier studies of these auctions tried to examine whether firms were prone to the bidder's curse. Capen, Clapp and Campbell (1971) claimed that winning firms earned be-
track. The estimated valuation of a track must be above this fixed cost for a firm to find it worthwhile to bid for a particular track. Therefore, bids will be observed only for those firms that have high estimates, leading to a selectivity bias. Correcting for this bias using Heckman's two step approach, Gilley and Karels (1981) find a negative relationship between bids and number of bidders.

It is difficult to interpret the above results. On the one hand, a negative, or inverted U shape, relationship between individual bids and the number of bidders is consistent with the predictions from the common value model. However, the alternative hypothesis is not clear. A failure to find this negative relationship could be an indication that 1) bidders are not behaving rationally and are falling prey to the winner's curse, 2) the auctioned good does not have a predominantly common value characteristic and another auction paradigm should be adopted, or 3) the game theoretic auction theory restrictions should be rejected (Laffont (1997)).

A weaker test of the theory is provided by an analysis of the relation between the winning bid and the number of bidders. For all auction mechanism the winning bid should be non-decreasing as the number of bidders increase. This is not very surprising. More bidders is tantamount to competitive pressure. Confirming this relationship with a particular data set should provide a general, albeit weak, test of auction theory.


See also Gaver and Zimmerman (1977) for another example of this line of research.
in all cases more bidders does increase the selling price, as theory predicts. They also attempt to test more specific predictions about the exact rate of growth of the winning bid within each bidding paradigm.

One of the most interesting studies using a reduced form approach is Hendricks and Porter (1988). They analyse the auctioning of offshore oil and gas drainage leases in the United States. In contrast to wildcat tracts, drainage tracts are those that are adjacent to places where deposits have already been found. This characteristic has important implications for the informational structure of the situation. In particular, Hendrick and Porter (1988) argue that when bidding for a tract, the firms that have the lease on neighbouring tracts will be better informed about the value of the tract than non-neighbouring firms. They adapt a common value auction model due to Engelbrecht-Wiggans, Milgrom and Weber (1983) where there is one informed bidder and a set of uninformed bidders. Because of the informational assumptions, the model makes strong predictions about the distribution of bids and the expected profits for each type of firm. They contrast these prediction with data on bids and profits of leases in the Gulf of Mexico auctioned between 1954 and 1969. They find that the data is consistent with the predictions of the asymmetric information model. The approach of Hendricks and Porter (1988) is, however, confined to situations where theory makes strong and testable predictions about equilibrium bidding behaviour. Unfortunately, in most auction situations this is not the case.

Finally, it is worth mentioning the numerous studies that analyse the impact of tendering for a good or service that was previously provided by some other mechanism. In many cases the good or service in question was supplied by a public entity, or monopoly firm, before tendering was introduced. These studies do not use information on the number of bidders but rather attempt to isolate the net effect of the change in policy. Usually they use regression analysis with a dummy variable for the type of selling mechanism. Domberger and Rimmer (1994) review these studies and find that there is
an indisputable negative effect of tendering on the cost of providing public services.

Although useful for some purposes, the reduced form models discussed above have several limitations. The most important is the impossibility of estimating underlying parameters that are needed to design an improved auction mechanism. For example, the distribution of underlying valuations is not recovered by this approach. Knowledge of this distribution is needed if the effects of changes in the bidding environment are to be predicted.

1.6.3 Structural econometric models of auctions

A structural estimation approach imposes the restrictions derived from game theoretic auction models in order to estimate all the underlying parameters of interest. The starting point for structural models is to assume that observed bids are the result of equilibrium strategies.

To illustrate the estimation methods we will concentrate on the symmetric independent private value environment. This environment is characterised as follows. There are $n$ risk neutral bidders for a single good. Each bidder has a valuation of $V_i$ for the good which is independent of the valuations of the other bidders (independent private value) and of any public information available on the good. Individuals do not know the valuation of their opponents. However, in deciding their bidding strategy they assume that the valuation of each of his opponents is distributed according to the continuous distribution function $F(v)$ with $v \in [u, \bar{v}]$. Furthermore, each individual assumes the same distribution for each of his opponents (symmetry) and this is common knowledge.

Depending on the selling mechanism (first-price sealed bid, Dutch descending auction, English ascending auction, or second-price sealed bid) there will be a symmetric equilibrium differentiable strategy for each player that maps a given valuation to a bid (Vickrey (1961); Riley and Samuelson
Formally,
\[ b_i = \sigma (v_i; F, n) \]  
where \( b_i \) is the bid tendered by individual \( i \), \( v_i \) is the value of the good for individual \( i \) and \( \sigma \) is the equilibrium bidding strategy function. The idea of structural approaches is to use data on bids and the bidding strategy (1.32) to estimate the distribution function \( F \).

Knowledge of the distribution \( F \) can then be used to study the effects of altering the selling mechanism. One popular application of structural models is to determine the minimum bidding level (reservation price) that an auctioneer should set to maximise the expected selling price of an object.

There are several methods that have been proposed to estimate the distribution of valuations. We will review them in turn\(^5\).

**Maximum Likelihood estimation**

The bid function (1.32) must be a strictly increasing function of valuations. Therefore, the bidding function can be inverted to obtain
\[ v_i = \sigma^{-1}(b_i; F, n). \]  
(1.33)

To estimate by maximum likelihood a parametric form for \( F \), say \( F(\cdot; \gamma) \), must be assumed, where \( \gamma \) is a vector of parameters of interest. Then the likelihood contribution of each observation can be computed. However, the likelihood contribution will depend on the auction mechanism used to sell the item and the nature of the data the researcher has at hand. We will begin with the case of an English ascending price auction and a second-price sealed bid auction. In general, we will assume that the researcher only observes the winning bid\(^5\).

---

\(^5\)See Hendricks and Paarsch (1995), Laffont and Vuong (1993), and Laffont (1997) for surveys on empirical analysis of auctions. Donald and Paarsch (1996) offer a complete treatment of the independent symmetric private values paradigm. They give the conditions under which \( F \) is identified from observed bids.

\(^5\)This is necessarily the case for a Dutch auction, where non-winning bids will never be observed. If all bids are observed, then the methods discussed below are still applicable.
In an English auction an auctioneer calls prices in ascending order until only one bidder is left. This bidder wins the auction and pays the price at which the auction stopped. The dominant strategy for a bidder is to stay in the auction until the price rises above his valuation. To see why this is so, notice that if the price rises above his valuation, and he happens to win, his utility will be negative, while if the price is below his valuation he can still expect a positive surplus from remaining active in the auction. Therefore, the equilibrium bidding strategy in an English auction is particularly simple,

\[ b_i = v. \]

The winning bidder will be the individual with the highest valuation for the good. He will pay a price equal to the second highest valuation, since that is the price at which there is only one bidder left and the auction stops. The likelihood of observing a winning bid of \( b^w \) is then the likelihood that \( b^w \) is the second highest valuation out of a sample of \( n \) valuations,

\[ L(b, n, \gamma) = n(n - 1)[1 - F(b^w; \gamma)]F(b^w; \gamma)^{n-2}f(b^w; \gamma). \]

Given a data set comprising information on the winning bid and the number of bidders for \( T \) independent English auctions (indexed by \( t \in \{1, 2, ..., T\} \)), the parameters \( \gamma \) can be estimated by maximising the log likelihood function,

\[ \ln L = \sum_{t=1}^{T} \ln L(b^w_t, n_t; \gamma). \]

It is straightforward to include some conditioning variables to account for observed heterogeneity of the good sold across auctions. Also, by suitable modifications of the likelihood for each winning bid, it is easy to extend the above approach to English auctions with reservation prices (see Hendricks and Paarsch (1995)).
In a second-price sealed bid auction (sometimes referred to as a Vickrey auction) the highest bidder is awarded the good but only has to pay a price equal to the second-highest bid. Just like the case of an English auction, the dominant strategy in this format is to bid one’s true valuation of the good. Bidding higher than the true valuation will only make a difference when the second highest bid is above this valuation. But then utility from winning will be negative. Likewise, bidding below one’s valuation will only make a difference when the second highest bid is between one’s bid and the true valuation. But then the bidder will lose the auction when it was profitable for him to bid higher and win.

Given that the equilibrium strategies are the same, the likelihood function for data from a second-price auction is exactly the same as in the English auction.

Paarsch (1989, 1991, 1997) has applied the above likelihood approach to the tendering of timber harvest rights in British Columbia. The sale of the harvest rights was sometimes undertaken by English type auction and other times by a first-price auction. For the data concerning English auctions, the method described above is easily implemented after making an assumption regarding the parametric family of the distribution of valuations. Once the parameters of the distribution is recovered, the researcher can use the results to simulate the effects of changes in the bidding rules. Paarsch uses his result to estimate the optimal reservation price that the auctioneer should set if he is to maximise expected revenues. It turns out that in his application the optimal reserve price is on average four times greater than the reserve prices used in practice.

The estimation of $F(v; \gamma)$ for English and second-price sealed bid auctions using likelihood methods is direct and computationally easy. For other auction mechanisms the approach is more complicated. The main reason being that the equilibrium bid function is no longer as simple as in the second-price or English auction. For example, in a first-price auction the highest bidder
wins the object and pays what he bid. The Bayesian Nash equilibrium bidding strategy\textsuperscript{56} for each bidder is given by

\[ b_i = v_i - \frac{\int_{v_i}^{\infty} F(\zeta; \gamma)^{n-1} d\zeta}{F(v_i; \gamma)^{n-1}}. \] \hspace{1cm} (1.34)

If we again denote the bidding strategy (1.34) as \( \sigma(v) \) then it is possible in principle to obtain the likelihood function\textsuperscript{57}. The probability that all bids are below \( b^w \) is the probability that all valuations (which are independent from each other) are below the level which makes bidders bid \( b^w \)

\[ Pr(b_i < b^w; i = 1, \ldots, n) = (F(\sigma^{-1}(b^w); \gamma))^n. \]

The density of a winning bid equal to \( b^w \) is then

\[ h(b^w; n, \gamma) = n(F(\sigma^{-1}(b^w); \gamma))^{n-1} f(\sigma^{-1}(b^w); \gamma) \frac{\partial \sigma^{-1}(b^w)}{\partial b}. \]

If the data contains information on the winning bid for \( T \) independent auctions, the log-likelihood function would be,

\[ \ln L = \sum_{t=1}^{T} \ln [h(b^w_t; n_t, \gamma)]. \] \hspace{1cm} (1.35)

Again, it is straightforward to include other observable conditioning variables to account for observable heterogeneity of the good sold across auctions. Also, the likelihood function need to be modified if there is a reservation price.

In a Dutch descending auction the selling price starts at a high value and gradually descends until a bidder stops the process. The winner is the individual who stopped the descent and pays the last posted price for the object. Strategically, the Dutch auction is equivalent to a first-price sealed bid auction. A bidders' decision problem is to calculate at what price it is optimal for him to stop the auction if someone else has not done so already.

\textsuperscript{56} Note that bidding behaviour is no longer in dominant strategies.

\textsuperscript{57} Notice that the equilibrium strategy \( \sigma(v) \) depends on \( F(v; \gamma) \) and the number of bidders \( n \). For ease of exposition we suppress this dependence in the notation.
Clearly, it must be at a value below his valuation if he is to earn a positive profit. It turns out that the price at which it is optimal to stop the auction, conditional on having a valuation \( v_i \), is given by equation (1.34). Therefore, the likelihood function is the same for a Dutch auction as in a first-price sealed bid auction\(^58\).

Unfortunately, the likelihood approach outlined above has several drawbacks in the case of a first price (or Dutch) auction. To implement this approach the equilibrium bidding function, \( \sigma(v) \), must be inverted. Even in the relatively simple case of a symmetric private values auction, inverting the bidding function (1.34) can be numerically demanding. When we relax some of the assumptions underlying the symmetric independent private values model, the computational difficulties become more severe. For example, in the case of an asymmetric private values auction (where \( F \) is not the same for all bidders) the Bayesian Nash equilibrium strategies are not solvable analytically, rather, they are expressed as a system of differential equations. Estimation by maximum likelihood would then require solving numerically the differential equation system for each of the \( T \) auctions at each trial parameter vector. Although feasible in principle, it would require large amounts of computer time\(^59\).

In addition, the maximum likelihood approach suffers from another, more severe, limitation. In the first price auction, the highest possible bid, \( \bar{b} = \sigma(\bar{v}) < \bar{v} \), will depend on \( \gamma \), the parameters of interest, since the equilibrium bidding function depends on \( F \). Thus, the range of possible values of bids will depend on the parameters of interest, which violates one of the regu-

\(^{58}\)Although theory states that both auctions are strategically equivalent, experimental studies seem to indicate that in practice bidders behave differently in both settings. See Milgrom (1989) and Smith (1991).

larity conditions needed to prove consistency and asymptotic normality of the maximum likelihood estimator. Donald and Paarsch (1996) propose a constrained maximum likelihood estimator. The likelihood function (1.35) is maximised subject to the following inequality constraints

\[ b_t^w \leq \bar{b}(n_t, \gamma) \quad \forall t. \quad (1.36) \]

The constraints force the observed winning bids to be within the range consistent with equilibrium bidding behaviour. Donald and Paarsch (1996) prove the consistency of the constrained maximum likelihood estimator. The asymptotic distribution of the estimator is quite involved and has only been derived when additional conditioning variables are discrete\textsuperscript{60}.

A third disadvantage of maximum likelihood in the context of auction data is that a parametric assumption must be made regarding the distribution \( F \). In some cases, such as an asymmetric auction, the relative shapes of the distributions is crucial to analyse the relative performance of different auction mechanisms (Maskin and Riley (1996). Therefore, assuming ex-ante a particular parametric form will preclude using the estimated model to analyse the relative merits of different auction mechanisms.

In spite of the disadvantages, the likelihood method for first price auctions has been applied by Paarsch (1989, 1991) to data on the British Columbia timber harvest rights. As mentioned before, these rights have been auctioned under both English and first-price mechanisms. Comparing the estimated results from each type of auction shows that the recovered distributions of underlying valuations differ.

Paarsch (1992) also uses maximum likelihood but because he compares estimates with the results from non-linear least squares we will defer the discussion of this paper until the next section.

\textsuperscript{60}Bajari (1997) circumvents these non-standard distribution problems by adopting a Bayesian approach. He assumes a prior distribution for the parameters of interest and then obtains a posterior distribution using the data.
Bajari (1997) also uses likelihood functions but within a Bayesian approach. This circumvents the asymptotic distribution problems raised above. Bajari (1997) assumes a prior distribution for the relevant parameters and obtains, using the sample data, the posterior distribution. Asymptotic distribution theory is not needed in this context.

**Non-linear least squares estimation**

An alternative to maximum likelihood is to use the conditional moments of the winning bid. To illustrate this approach, assume for the moment that we could obtain an expression for the first moment of the winning bid, \( m(n_t, \gamma) = E[b^w] \). Then a non-linear least squares estimator would be the vector of parameters that minimised the sum of squared residuals,

\[
\min_{\gamma} \sum_{t=1}^{T} (b^w_t - m(n_t, \gamma))^2. \tag{1.37}
\]

To gain efficiency higher moments of the winning bid could also be used\(^6\). The difficulty with the least squares method lies in obtaining \( m(n_t, \gamma) \). One alternative is to assume some parametric family for \( F \) such that this moment can be easily derived. This was the alternative adopted by Paarsch (1992).

Paarsch (1992) is interested in testing whether the common value or the private value paradigm applies to the tendering of tree planting contracts in British Columbia. In order to achieve this, he estimates several models under each informational assumption both by maximum likelihood and by non-linear least squares. Because of the distributional assumptions made, Paarsch (1992) is able to obtain closed form expressions for the first two moments and the likelihood function of almost all models specified. He then contrasts the non-linear least squares parameter estimates with the maximum likelihood parameter estimates using a Hausman type specification test for each model (Hausman (1978)). Notice that Paarsch (1992) does not compare

---

\(^6\)See Laffont and Vuong (1993).
results from different models using some non-nested test procedure but rather compares estimated consistency within each paradigm. He finds that the common value auction (and specifically Rothkopf's (1969) specification) is not rejected by the data while all of the private value specifications performed poorly.

Laffont, Ossard and Vuong (1995) propose a second alternative for computing \( m(n_t, \gamma) \). They exploit the revenue equivalence theorem to obtain this moment. This theorem states that the expected winning bid in a first-price symmetric private value auction is equal to the expected winning bid in a second price symmetric private value auction (Vickrey (1961); Riley and Samuelson (1981), Myerson (1981)). As discussed above, it is a dominant strategy in a second-price auction to bid the true valuation. Therefore, the expected winning bid in a first-price auction will be equal to the expected winning bid in a second-price auction, which in turn is equal to the expectation of the second highest valuation in a sample of \( n \) bidders,

\[
m(n_t, \gamma) = E[v_{(2:n)}] = \int_{v} v(n_t(n_t - 1))(F(v; \gamma))^{n_t - 2} (1 - F(v; \gamma))f(v; \gamma)dv \tag{1.38}
\]

where \( v_{(2:n)} \) denotes the second highest valuation out of a sample of \( n \) valuations. Notice that the equilibrium bidding strategy \( \sigma(v) \) or its inverse is no longer required.

One final difficulty is the fact that, save for some particular distributions, \( m(n_t, \gamma) \) will contain integrals and therefore needs to be evaluated numerically. To circumvent computational difficulties, Laffont, et al (1995) use an unbiased simulated estimator analogous to methods proposed by McFadden (1989) and Pakes and Pollard (1989). They draw \( J \) independent samples of \( n_t \) valuations from \( F(v; \gamma) \) and use the average of the second highest valuation
as a simulated estimator of the expected winning bid. Formally,

\[ \tilde{m}(n_t, \gamma) = \frac{1}{J} \sum_{j=1}^{J} s_{tj} \]

where \( s_{tj} \) is the second highest valuation of sample \( j \). Laffont, et al (1995) then show that the ensuing “simulated non-linear least squares estimator” (SNLLS) is consistent and asymptotically normal provided that an additional term is added to the objective function to correct for simulation bias. This bias occurs because, although the simulated estimator \( \tilde{m}(n_t, \gamma) \) is unbiased, the objective function is not linear in this simulated moment. In particular, the estimator given by

\[
\min_{\gamma} \frac{1}{T} \sum_{t=1}^{T} \left( (b_t^w - \tilde{m}(n_t, \gamma))^2 - \frac{1}{J(J-1)} \sum_{j=1}^{J} (s_{tj} - \tilde{m}(n_t, \gamma))^2 \right)
\]

is consistent and asymptotically normal. As an illustration, Laffont, et al. (1995) apply this method to the aubergine market in Marmande, southern France. They assume a log-normal distribution for bid valuations and use 20 draws for the simulations.

The non-linear least squares estimator still has some drawbacks. The non-simulated version will be computationally demanding when the moment, \( m(n_t, \gamma) \), is difficult to evaluate. In many cases, including asymmetric auctions, the equilibrium bidding functions are expressed as differential equations. The moments would then have to be computed numerically for each auction and each trial of \( \gamma \). This approach could be quite demanding. For example, Elyakime, et al. (1995) report that to estimate their model, which required solving one differential equation numerically, using NLLS took 100 hours of CPU time on a SUN network. They had data on 281 bids for 70 auctions and used only one additional exogenous conditioning variable.

The differential equation system describing equilibrium bid functions in an asymmetric auction would make NLLS estimation even more demanding\(^{63}\).

The simulated NLLS estimator proposed by Laffont, et al. (1995) is an ingenious method of avoiding the above computational problems. Unfortunately, it can only be used when the revenue equivalence theorem holds between a first-price auction and a second-price auction. This rules out most bidding environments except for the symmetric independent private value paradigm (Milgrom and Weber (1982)).

Finally, to implement either version of the NLLS estimator, a parametric assumption must be made about \(F\). As was discussed above, making parametric assumptions might be inconvenient in cases where the shape of the distribution of valuations is crucial to determine the relative performance of different auction mechanisms.

**Non-parametric estimation**

Elyakime, et al. (1994) have proposed a non-parametric estimation method that is applicable to a wide range of auctions. To explain their method, it is necessary to begin with the first order condition for the Bayesian Nash equilibrium of the symmetric private value auction. The first order condition is the differential equation

\[
\frac{\partial \sigma(v)}{\partial v} = (n - 1)(v - \sigma(v)) \frac{f(v)}{F(v)}
\]

(1.39)

of which the Bayesian Nash equilibrium strategy (1.34) is a solution when the boundary condition is \(\sigma(v) = v\).

Elyakime, et al. (1994) work directly with the differential equation (1.39). They exploit the relation between the distribution and density of bids and the distribution and density of valuations. In particular, if we define \(G(b)\)

\(^{63}\)See Marshall, et al (1994) for a numerical algorithm to solve for the equilibrium bidding functions of an asymmetric auction when valuation distributions are uniform.
and \( g(b) \) as the distribution and density of bids, respectively, we have the following relationships

\[
G(b) = F(\sigma^{-1}(b)) \quad (1.40)
\]

\[
g(b) = f(\sigma^{-1}(b)) \frac{1}{\frac{\partial \sigma(v)}{\partial v}}. \quad (1.41)
\]

Substituting into (1.39) results in

\[
v = b + \frac{G(b)}{(n-1)g(b)}. \quad (1.42)
\]

Equation (1.42) states that, if bids are the result of equilibrium strategies, then valuations can be expressed as a function of bids and the distribution and density function of bids. This allows a simple yet powerful estimation strategy.

Guerre, Perrigne and Vuong (1995a), following Elyakime, et al. (1994), propose a two-stage non-parametric estimation technique to recover the underlying distribution \( F \) of valuations. In the first stage, use the bid data to estimate \( G(\cdot) \) and \( g(\cdot) \) non-parametrically. Use the resulting estimates, say \( \tilde{G}(\cdot) \) and \( \tilde{g}(\cdot) \) to generate for each bid in the data set a pseudo-valuation using (1.42), \( \tilde{v}_t \). Then use the pseudo-sample of valuations to estimate \( F(\cdot) \) non-parametrically.

A necessary and sufficient condition for \( F(\cdot) \) to be identified is that the valuations, as expressed by equation (1.42), be increasing and differentiable in \( b \). The statistical and asymptotic properties of this two-stage estimator can be found in Guerre, et al. (1995a; 1995b).

The estimator described above requires knowledge of all the bids, not just the winning bid. To estimate \( G(b) \) above, all bids need to be used. However, Guerre, et al. (1995a) extend the method to the case where only the winning

---

64This is Theorem 1 in Guerre, et al. (1995a) which is proved in Guerre, et al. (1995b). For the case of asymmetric bidders the same result holds (see Proposition 2 in Elyakime, et al. (1994)).
bid is observed (in a symmetric context) and to the case where the auction includes a reservation price.

Given the small samples that are usually available for auction data, and given that non-parametric estimation is less efficient than the parametric counterpart, the small sample properties of the non-parametric estimator must be considered. Monte Carlo evidence presented in Guerre, et al. (1995a) shows that the estimator is precise even for small samples of 250 bids. This result, however, is only valid for the symmetric private value auction and may not carry over to more demanding auction situations.

Elyakime, et al. (1994) apply this non-parametric estimator in their study of timber sales in southwest France. The particular auction and bidding equilibrium they analyse is more complicated than the symmetric model presented above. In the timber auctions, the seller submits a secret reservation price for the lots and bidders (sawmill owners) do not know this value before they submit their bids. Elyakime, et al. (1994) present the features of this first price auction with secret reservation prices and characterise the equilibrium strategy of the seller and the (symmetric) buyers. The data consists of a total of 281 bids for 70 lots. Using kernel estimators, they obtain estimates of the distribution of bids for all agents. Using analogous equations to (1.42) they generate a pseudo-sample of valuations and estimate their distribution. The non-parametric results show that valuations are approximately log-normal. Therefore, they then postulate this parametric family for the distribution of valuations and use a non-linear least squares estimator to re-estimate the model. The parametric results are used to calculate the potential gains to the seller of announcing the reservation price before the auction rather than maintaining it secret.

The advantage of the approach presented in this section is that it can be applied to a wide set of auction situations. In particular, it is well suited

---

65 Another example of the use of the non-parametric estimator is Elyakime, Laffont, Loisel and Vuong (1995)
to analyse asymmetric auctions. As will be seen below, in an asymmetric context bidding strategies are expressed as a system of differential equations which are not solvable analytically. Computational problems become excessive if a likelihood approach or a non-linear least squares estimator is used. In contrast, equations similar to (1.42) result in the asymmetric case. Notice that there is no need to compute the equilibrium bidding strategy in the non-parametric approach. Numerically, this strategy can be found by plotting bids against their corresponding pseudo-values. Furthermore, the non-parametric approach is especially attractive for applications to asymmetric auctions. In this case, the relative merits of different auction mechanisms depend on the shape of the distribution of valuations for different bidders. Therefore, one would like to avoid making a-priori parametric assumptions in this respect.

Other estimation approaches

Deltas (1996) proposes an alternative estimation method to those presented above. He estimated a structural model for the tendering of road construction contracts in Connecticut. His data set contains information on all the bids for a number of auctions, not just the winning bid. Since firms bid for a contract, the underlying valuations are the costs for each firm and the bids are the price they are willing to receive to undertake the contract.

The first step in his approach is to estimate the distribution of costs, $F(c; \gamma)$. Imposing the same parametric family on all the observed contracts, an estimate $\hat{\gamma}_t$ is obtained for each contract using all the bids received for that particular contract. $\hat{\gamma}_t$ is estimated by maximum likelihood and a method of moments estimator. The estimation stage is undertaken for several assumed parametric families for the cost distribution. Using specification tests, the researcher is able to determine which parametric family is consistent with the data.

Deltas (1996) assumes three possible distributions for his study of the
tendering of construction contracts in Connecticut, the uniform, exponential and log-normal. For two of these distributions, the bidding functions in a symmetric private values setting have closed form solutions which facilitates the estimation procedure. For the log-normal distribution he uses numerical techniques. The fact that all bids are observed simplifies the estimation somewhat since the density of each bid is not related to order statistics.

The second stage of the estimation procedure is to relate the estimated parameters, \( \hat{\gamma}_t \), to project characteristics. By assuming a functional relationship between \( \gamma \) and project characteristics, and using the statistical properties of the estimator \( \hat{\gamma}_t \), it is possible to analyse the impact of characteristics on the heterogeneity of costs across auctions. Furthermore, it is possible to take into account unobserved heterogeneity as well as observed characteristics that vary from auction to auction.

Deltas (1996) approach has several advantages. By breaking up the problem into several stages it avoids potentially complicated computational problems in estimating a full blown structural model. In addition, it allows for testing for the parametric family of cost distributions without specifying the relationship between the parameters of the distribution and the characteristics of the good being auctioned. Once the parametric form for the cost distribution has been determined there is no need to repeat this stage. Different specifications linking \( \hat{\gamma}_t \) to project characteristics can be analysed without having to re-estimate the computationally intensive first stage. When new data becomes available there is also no need to re-estimate the first stage for the old data. The drawback from Deltas (1996) method is that some efficiency is lost by performing the estimation in stages. Also, the method assumes that information on all bids, not just the winning bid, is available. This is often not the case with auction data.
1.6.4 Asymmetric auctions and the tendering of refuse collection contracts in England

The third essay in this thesis deals with the tendering of refuse collection contracts by local authorities in England. Several issues regarding these auctions are examined using a novel data set that includes information on the number of tenders received for each contract.

Since 1988, local authorities in the United Kingdom have been obliged by law to tender their contracts for blue collar services, including refuse collection. This has been a politically divisive issue since its introduction, and its future is now uncertain given the labour government’s election pledge to abolish it.

Chapter 4 consists of two parts. In the first, reduced form analysis is undertaken to examine the relation between costs and the number of bidders. As the review of section 1.6.2 illustrates, there is a small literature related to the effects of the number of bidders on the sale price of an item. Theory would predict that as the number of bidders increases, competitive pressures will reduce the winning bid for a contract. Confirming this proposition with data constitutes a weak test for auction theory.

In the context of refuse collection contracts in England, showing that a negative relationship exists between refuse collection costs and the number of bidders has important policy implications. It shows that contract tenders are functioning as expected and that they provide a useful tool for extracting rents from natural monopoly providers. In addition, it questions the judiciousness of plans to replace tendering by another regulatory instrument.

The results of chapter 4 clearly show that more bidders lowers the expenditure on refuse collection by local authorities. Moreover, this conclusion still holds when allowance is made for potential endogeneity problems regarding the number of bids variable.

The second part of chapter 4 estimates a structural model of refuse collec-
tion contract tendering. The motivation for the analysis is the observation that a significant proportion of the contracts have been won by the local authorities' Direct Service Organisations (DSOs) instead of the private contractors that compete with them. A DSO is the in-house team that provided the service in a local authority before competitive tendering was introduced. The high success rate of the DSOs implies an asymmetry in the bidding process and may be a sign that local authorities are somehow biasing the tendering process in favour of the DSO. Structural analysis may shed some light on this issue.

Using the non-parametric approach suggested in Elyakime, et al (1994) we recover the underlying cost distributions assuming that there are two types of agents, DSOs and private firms. Because it is assumed that the auction is asymmetric, the equilibrium bidding strategies of agents are expressed as differential equations that have to be solved numerically. In this context, the simplicity of the approach proposed by Elyakime, et al (1994) make it a natural estimating strategy.

The structural estimation undertaken in chapter 4 adds another example to the nascent econometrics of auctions literature. It is also one of the few application to an asymmetric first-price auction context.

Comparing the estimated cost distributions for each type of firm suggests that DSO costs are lower than private firms. Therefore, the high success rate could be attributed to an incumbency cost advantage of DSOs, or to any factor that would lower DSO costs vis-à-vis private firms costs. One explanation that is not inconsistent with the estimation results is that local authorities are biasing the tendering process in favour of the DSO. For example, local authorities may be designing contracts that serve to increase private costs relative to DSO costs. They may also be imposing certain requirements on private firms, such as the posting of performance bonds, which serve to increase their costs.
1.7 Conclusions

The new theory of regulation emphasises the importance of unobservable latent variables for the optimal design of regulatory mechanisms. On the resource allocation front, unobservable variables usually represent taste heterogeneity among individuals or households. Furthermore, this taste heterogeneity coupled with non-linear prices will result in some kind of sorting behaviour by consumers according to their taste parameter. This will have non-trivial implications for empirical analysis, as will be seen in the next chapter.

On the production side of the regulatory model, the unobservable variables represent influences on costs that are crucial to determine the optimal cost reduction effort that a regulated firm should exert. The unobservable variable in a production context also has econometric implications (Wolak (1994)).

In both the demand and supply side of a natural monopoly market, empirical analysis must recognise the problems associated with unobservables if consistent parameter estimates are to be obtained. However, the need for consistency in parameter estimates is not the only reason for paying close attention to unobservable latent variables. In many cases, knowledge of these latent processes is needed to answer many of the most interesting questions faced by regulators. Issues related to the design of optimal tariffs or the impact of changes in the regulatory environment can only be addressed if the distribution of unobservables is known. This calls for estimation techniques aimed directly at recovering these underlying distributions. Structural modelling, as opposed to reduce form approaches, are therefore unavoidable in this context.

Structural models have the disadvantage that they may be prone to misspecification. As the discussion in chapter 3 reveals, sometimes many arbitrary parametric assumptions must be made in order to identify all the
parameters of interest. However, there seems to be no alternative to structural modelling if quantitative answers are to be given to questions regarding changes in the regulatory environment.
Part I

Regulation: Demand issues
Chapter 2

Energy Demand Modelling and Tariff Rebalancing in the United Kingdom Gas Market

2.1 Introduction

The domestic energy markets in the United Kingdom are still in a process of structural change. Since the privatisation of British Gas in 1986, competition in gas supply has been introduced in steps. Initially, only the industrial market (premises consuming more than 25,000 therms per year) was liberalised and, later, competition in this segment of the market was actively encouraged by OFFERS. Starting in 1992, competitors to British Gas were allowed to supply the 2,500 therm to 25,000 therm market. In contrast, the market segment defined by premises with less than 2,500 therms of gas consumption per year has been serviced exclusively by British Gas since privatisation. This segment is composed mostly of domestic households and has been regulated by an RPI-X formula in order to check British Gas' monopoly power\(^1\). In April, 1996, limited competition for the supply of household domestic gas

\(^1\)For a review of the privatisation of British Gas and the subsequent regulation of the gas market see Vickers and Yarrow (1988) and Armstrong, et al.(1994). In 1997, British Gas will be demerged into two companies, Centrica, which includes the gas supply arm of the old British Gas, and Transco, the natural monopoly transportation business. See Waddams-Price (1997) for a recount of recent events in the gas industry.
was introduced, with full scale competition expected to develop in the next few years. Competition for the supply of electricity to households is expected to begin in 1998.

The introduction of competition in the supply of these energy goods will force tariffs to become more cost reflective. Until now, maintaining cross subsidies between consumer groups has not posed any difficulties, given the monopoly concessions enjoyed by electricity suppliers and British Gas. Profits lost by subsidising one group of consumers is compensated by high profit margins for other groups. Competition is likely to change this. New entrants will target market segments where current prices are above supply costs. They will have scant incentives to supply groups where tariffs are below costs. To survive, incumbents -who have universal services obligations- will be forced to end internal cross subsidies.

There is a presumption that one of the most important cross subsidies present in the current pricing structure for energy goods is the balance between the standing charge (which is independent of the amount of energy consumed) and the unit price of energy. The fixed costs of supplying a household -in terms of metering and billing expenses- are above the current standing charge. On the other hand, the price for each unit of energy is above the supply costs. As competition takes hold in these markets we would expect a rebalancing to occur between these two components of energy prices\(^2\).

The rebalancing of the fixed and variable charge might be justified on efficiency grounds -or at least unavoidable in a competitive market- but the distributional consequences could be a cause for concern. Energy goods are necessities and as such they represent a higher proportion of the expenditure of low income households. This partly explains why domestic energy price changes are so contentious politically. Witness, for example, the special tax

\(^2\)However, since the gas transport industry will remain a regulated monopoly, some cross-subsidies in final competitive price may be sustained if the regulator sets the intermediate transport prices accordingly.
treatment accorded to domestic energy goods in the United Kingdom. These goods are not subject to the full Value Added Tax\(^3\). The indisputable fact is that, unless compensated by off-setting welfare policies, increases in fuel prices have regressive distributional consequences.

Since low income households purchase lower than average amounts of energy, the standing charge represents a higher proportion of their total energy bill. Thus, tariff rebalancing resulting in a higher fixed charge is more likely to have a negative impact on low income households. The savings due to the lower unit price will be applied to a smaller level of energy consumption and will not compensate for the increase in the standing charge.

This adverse distributional impact may explain why regulators in these industries are reluctant to allow an increase in the fixed charge\(^4\). In the case of gas, there has been a supplementary price cap applied to the standing charge. This is in addition to the well known RPI-X cap applied to gas revenues as a whole. Also, it seems that OFGAS has implicitly sanctioned the preservation of some level of cross subsidy in gas prices by setting the fixed charge for Transco (the gas transporting company) below what British Gas and competitors expected (Hancock and Waddams-Price (1995)).

The relation between energy prices and welfare has been analysed in Hancock and Waddams-Price (1995) and Burns, Crawford and Dilnot (1995). These studies confirm that poorer households are disproportionately affected by increases in energy prices. Unlike these previous studies, in this chapter we use an econometric model of household consumption behaviour to examine these issues. Thus, our model allows for behavioural responses to price

\(^3\)The Value Added Tax on domestic energy goods was raised from 0% to 8% in April, 1994. However, the intention of the government to raise it to the full 17.5% in April, 1995, was struck down in Parliament. The new Labour government plans to reduce VAT on fuels to 5%.

\(^4\)There are other reasons why regulators might not wish to grant firms unlimited freedom to set their tariff structures. If regulated by an average revenue cap, firms may have an incentive to "over-balance" their tariffs by raising the fixed charge above what is socially convenient (see Armstrong, et al (1995) and the discussion in section ?? of chapter 1).
changes. This should yield more precise and quantitative welfare results than those in previous studies.

An energy demand model is estimated using Family Expenditure Survey (FES) data for the United Kingdom (excluding Northern Ireland). This estimation considers, as far as the data make it possible, the non-linear nature of energy prices and the discrete choice connection decisions faced by consumers.

The demand model is then used to calculate the welfare impact of estimated tariff rebalancing in the domestic gas market. In particular, published information on the possible magnitude of the current cross subsidy in gas prices is used to construct several scenarios of tariff changes. We then examine the welfare consequences of each scenario.

This work is similar in spirit to Wolak (1994) who uses similar techniques to analyse the impact of tariff restructuring in the United States telecommunications market. He presents evidence that an increase in domestic telephone tariffs (in tandem with a reduction in long distance prices) will probably be welfare improving.

Our results would seem to suggest that the misgivings of consumer groups and the regulator with respect to tariff rebalancing may be justified. Although the majority of households would gain from tariff rebalancing, there is a significant group that would be made worse off. Furthermore, the latter group is composed of poorer and more vulnerable households. We also show that if competition reduces customer related costs by between 20% and 30%, and these costs reductions are translated into lower fixed charges, then almost all households stand to gain. However, if the potential cost savings are below this 20% to 30% range, then tariff rebalancing sparked by the introduction of competition does pose distributional conflicts.

\footnote{For earlier studies of energy demand using FES data see Baker and Micklewright (1987) and Baker, Blundell and Micklewright (1989).}

\footnote{See Chapter 1 for a review of this paper.}
The analysis of the welfare impacts of tariff changes is important in its own right. Results can be used to evaluate the need for special programs to alleviate negative distributional impacts and to help design these programs if required. But, welfare analysis may also have more direct policy implications for regulators. In some cases, the threat of adverse welfare consequences may be used by a monopoly incumbent as an argument against proposed regulatory reforms\(^7\). This issue has played a part in the discussions surrounding reforms of the telecommunications market in the US (see Kaserman, David, Mayo and Flynn (1990)). In the UK gas market, the potential welfare consequences of tariff rebalancing has also been an issue in the introduction of competition (see The Observer (1996)). Analysing the veracity of distributional and welfare claims may, therefore, be a vital input to regulatory reform in utility industries.

2.2 Tariffs, price discrimination and the tax benefit system

Varian (1989) defines price discrimination as a situation where the ratio of prices of two similar goods is different from the ratio of their marginal costs. A cross-subsidy, on the other hand, refers to a pricing structure by a multi-product firm such that entry by a rival in one or more sub-markets is profitable (Faulhaber (1975), Baumol, Panzar and Willig (1982), and Curien (1991)). If there are no economies of scope, and profits are zero, then both concepts coincide. We will assume this is the case in the following discussion, so both terms will be used interchangeably.

It is believed that domestic gas tariffs are discriminatory and possibly result in cross-subsidies between consumer groups. There is a presumption of price discrimination in four dimensions\(^8\):

\(^7\)Laffont and Tirole (1993) note that firms may have a strategic incentive to cross subsidise in order to make it politically more difficult for a regulator to introduce reforms.

\(^8\)House of Commons (1994).
i) Between consumers in different regions. The supply cost is believed to differ between the regions in the east, close to the landing sites of the North Sea gas, and those further west, which require costly transportation for supply. Currently the tariff structure is uniform for the whole country, implying price discrimination in favour of western regions to the detriment of consumers in eastern regions. Tariffs in the industrial market for pipeline transportation services vary according to distance transported, which is inconsistent with the current homogeneous tariffs faced by domestic customers.

ii) Between large and small customers. According to British Gas the balance between the fixed charge and variable prices in gas tariffs favours low consumption households. The fixed costs of supply is above the current fixed charge while the variable cost is below the variable price.

iii) Between customers with an even load factor and those with seasonally varying demands. Since customers with varying demands are responsible for costly peaking storage facilities, the cost of supplying those customers is higher. This might call for a higher tariff for households that use gas central heating, for example, and thus have a more seasonally varying demand than those that use gas exclusively for cooking and water heating.

iv) Between good payers and bad payers. The administrative costs attributed to households that are slow to pay or do not pay are significant. According to the report by the Trade and Industry Committee, British Gas regards this as the most important source of price discrimination.

In this paper we will concentrate on ii) above. Regional cost differences do not seem to be important in magnitude, reaching between 2%-4% of average costs (House of Commons (1994)). Therefore, the unbundling of regional cross subsidies does not seem to pose alarming welfare consequences. Lack of information prevents us from analysing iii) and, further, it does not seem to

There could also be intra regional price discrimination between rural and urban consumers. However, the evidence presented to the House of Commons indicates that this does not seem to be an issue. See House of Commons (1994).
be a policy issue regarding the introduction of competition. A comment regarding iv) is warranted. There is already a tendency for British Gas to offer lower prices for customers paying by direct debit from their bank account. It is possible that with the introduction of competition the price reduction for prompt payers will intensify. The effect of this tendency, however, will not be analysed here\footnote{See Hancock and Waddams-Price (1995) for more on this issue.}.

\subsection*{2.2.1 Optimal Tariffs}

In chapter 1 the optimal tariff literature was reviewed. One conclusion from that discussion was that it might be optimal from a social perspective to have tariffs that do not reflect the underlying cost structure.

It is a well known proposition in economics that prices should be equal to marginal costs. It guarantees that the social value of the last unit produced is equal to the social cost of producing it. This still holds in natural monopoly industries where increasing returns to scale imply that it is economical to have only one producer. However, in these industries the revenue raised by marginal cost pricing will not recover the full costs of production and the firm will be making a loss. The shortfall in revenues could be recouped through a direct transfer from the government, as in the regulatory model of chapter 1. However, to make a transfer, the government must raise revenue from other sectors of the economy. The distortionary taxation needed to raise this revenue will impose welfare costs on these other sectors. Therefore, it might be preferable for a government to fund the deficit of a natural monopoly by setting prices above marginal costs rather than make a transfer using general tax revenues.

In many cases, regulators are not allowed to make transfers to regulated firms. In this case, tariffs must be set so that revenues cover all costs, which also forces a departure from marginal cost pricing.
What is the optimal departure from marginal cost pricing if the firm is to be self-financing or if revenues to fund transfers create other distortions in the economy? This was the topic of section 1.4. We will briefly review those results.

If the firm is constrained to charge linear prices, then these prices should be set above marginal costs, according to the "Ramsey" equation (1.21). However, there is an inefficiency involved in this pricing rule since the quantity sold will be lower than what is socially optimal.

Two-part tariffs will usually be better than linear prices. The advantage of a two-part tariff is that it allows the marginal price to be set closer to marginal supply cost and any shortfall in revenue is recovered through a fixed charge that is independent of the quantity demanded.

There are two problems with this type of tariff. First, some consumers may avoid paying the fixed charge by not participating in the market. This will be the case if the surplus from consuming is not large enough to compensate for the fixed charge. This, in spite of the fact that they would purchase positive amounts when faced with a linear price. There is a trade-off then between raising revenue by setting a higher fixed charge, and lowering market penetration, or setting a higher marginal price and driving a wedge between price and marginal cost.

The other consideration for the setting of a two-part tariff is distributinal. The balance between the fixed price and the variable price will determine the proportion of the revenue shortfall paid by different consumers. A high marginal price makes the burden of the budget deficit fall on the high purchasers. If the welfare weight of individuals with small consumption levels is higher than the welfare weight of individuals with a high consumption level (as would be the case with normal goods and a welfare function that weighs individuals inversely to their marginal utility of income), then it is welfare improving to raise the marginal price and lower the fixed charge.

The trade-offs discussed above imply that there is an optimal two-part
tariff structure for a particular situation. Ignoring distributional effects, the optimal mark-up derived in chapter 1 was

\[ \frac{p - \frac{\partial C}{\partial q}}{p} = \frac{\lambda}{(1 + \lambda)} \left[ \frac{1}{\tilde{\eta}} \right] \left( 1 - \frac{q^*(\tilde{\beta})}{\bar{q}} \right) \]  \hspace{1cm} (2.1)

and

\[ a - \left( \frac{\partial C}{\partial N} - (p - \frac{\partial C}{\partial q})q^*(\tilde{\beta}) \right) = \frac{\lambda}{(1 + \lambda)} \frac{1}{\tilde{\eta}_N} \]  \hspace{1cm} (2.2)

where \( p \) is the variable unit price for the regulated good, \( a \) is the fixed charge, \( \frac{\partial C}{\partial q} \) is marginal supply cost, \( \frac{\partial C}{\partial N} \) is the marginal consumer related costs of supplying an extra household, \( q^*(\tilde{\beta}) \) is the consumption level of a consumer (type \( \tilde{\beta} \)) who is indifferent from purchasing positive amounts of the good and not purchasing at all, \( \lambda \) is the cost of public funds, \( \tilde{\eta} \) is the price elasticity of demand holding \( \tilde{\beta} \) constant, and \( \tilde{\eta}_N \) is the elasticity of the number of consumers who purchase positive amounts of the good with respect to the fixed charge\(^{11}\).

As in the case of linear prices, equations (2.1) and (2.2) show that it may be socially optimal to set price-cost margins differently for the fixed charge and the variable price. This will depend on the magnitude of the relevant elasticities. The point is that it is possible that the socially optimal tariffs will not be cost reflective.

The above pricing results were derived assuming that a single monopoly firm operated in the industry. The domestic gas supply market in the United Kingdom is, however, evolving towards a competitive industry. How relevant are the above optimal pricing results in this context?

First, one might argue that the incumbent will remain the dominant firm in the industry for the foreseeable future. It will, therefore, remain regulated.

\(^{11}\)The left hand side of equation (2.2) is also a price cost mark-up. The cost of supplying an extra consumer is the customer related costs minus the profits that will be made from the purchases of the new consumer. Therefore, both pricing conditions can be regarded as Ramsey pricing formulas.
and the optimality of its pricing structure should be evaluated. To this end, the above results still hold. However, the demand function must be redefined as the residual demand faced by the incumbent. Furthermore, since this company will be regulated by an RPI-X formula, it is important to evaluate the incentives that this regulatory scheme gives to the firm to set tariffs in line with the socially optimal levels. This line of research has been pursued by Armstrong and Vickers (1991) and Armstrong, et al. (1995).

Second, there is the issue of the optimal pricing structure for gas transport. It is believed that the only natural monopoly part of the gas supply industry is transport and storage. The transportation business will remain subject to price regulation. The point for the current discussion is that by setting the access conditions the regulator will be able to control indirectly the pricing structure for delivered gas. Therefore, the optimal monopoly pricing results are still relevant but only for the transportation side of the industry. In fact, it seems that the regulator has already imposed a cross subsidy by setting the standing charge for British Gas' transportation services below what British Gas and independent suppliers had suggested would reflect the true cost structure\(^{12}\).

### 2.2.2 Tariffs and the tax-benefit system

Another criticism of the relevance of the optimal pricing literature is the relation between optimal tariffs and the tax-benefit system. The analysis of chapter 1 introduced a cost of public funds to take account of the distortionary effects of taxation. The general equilibrium justification for the cost of public funds approach relies on the existence of an imperfect tax system (Laffont and Tirole (1993)). However, it is a well known result in the optimal taxing literature that under certain conditions – principally that leisure be weakly separable from consumption in the utility function and that the

\(^{12}\)House of Commons (1994).
government is applying an optimal income tax-benefit policy—optimal indi
direct taxes should be uniform (Atkinson and Stiglitz (1976)). This result
implies that indirect taxation should not be used as a resource allocation or
distributive tool. In the context of utility pricing, it would dictate that the
price mark-up for the fixed and variable charge for the energy goods should
be similar to any other good in the economy, namely the Value Added Tax
rate. The possible disconnection that this pricing structure would induce, as
well as any distributional consequences, would be addressed through direct
transfers (or a lower overall VAT rate) to households.

In summary, there are two views to optimal tariffs. If the tax and benefit
system is imperfect and policies that redress the negative impacts of tariff
changes cannot be implemented, then it may be socially convenient to allow
price mark-ups and cross subsidies to be determined by the demand behav­
iour of consumers and distributional considerations. On the other hand, if
a sophisticated tax and benefit system is in place, then the case for cross
subsidies and distortionary mark-ups is weaker. As Vickers notes:

"Should regulators take distributional considerations into account
in making their decisions? I think not—except perhaps where spe­
cific duties with distributional aspects, together with the financial
and other means of carrying out those duties, have been given to
them by government or parliament. The basic reason is that dis­
brutional policies are generally better pursued by other branches
of government, which is not to say that the latter can achieve a
distributional outcome that could not be improved by the use of
regulatory instruments. Rather, the view is that the advantages
of regulators having discretion to pursue distributional ends are
outweighed by disadvantages of capture, influence activities, un­
certainty, and unaccountability. Regulators, perhaps like central
bankers, should have focused objectives.”\textsuperscript{13}

However, even if one agrees with the above comment, an energy demand model is still useful to estimate the welfare effects of tariff changes and thus contribute to the design of the optimal tax and benefit policies. This is the approach taken in this chapter. We now turn to the demand model.

\section{2.3 Household energy demand model}

Domestic energy goods in the United Kingdom (gas and electricity) are sold using two-part tariffs\textsuperscript{14}. The interesting aspect of multi-part prices is that they produce kinks and/or discontinuities in the budget constraint. This is important from an econometric point of view given that consumer’s optimal response is a discontinuous function of the parameters being estimated. Econometric techniques have been developed for dealing with non-linear budget constraints and non-consumption in demand systems\textsuperscript{15}. In our estimation strategy we will take account of the econometric difficulties related to two-part tariffs.

\subsection{2.3.1 The QUAIDS demand system}

In this chapter, household preferences are represented by the QUAIDS utility function of Banks, Blundell and Lewbel (1997). This utility function is flexible enough to portray a wide range of consumption patterns and elasticities, while still being consistent with consumer demand theory. It has the particular feature that the derived share equations for goods are quadratic

\textsuperscript{13}Vickers (1997).

\textsuperscript{14}In reality, they are sold using a menu of multi-part tariffs. For example, there is the standard two-part tariff for credit customers, but a different two-part tariff for pre-payment meter customers. In the case of electricity, there is also a special night-time rate for consumers that opt to consume electricity mainly at night.

functions of the logarithm of real expenditure. It nests two other popular demand systems, the Almost Ideal Demand System of Deaton and Muellbauer (1980) and the Translog model of Jorgenson, Lau and Stoker (1982).

Indirect household utility is given by

$$\ln(V) = \left( \frac{\ln(m) - \ln(a(p))}{b(p)} + \varphi(p) \right)^{-1}$$

where $\ln(V)$ is the indirect utility, $\ln(m)$ is the logarithm of expenditure, $a(p)$ and $b(p)$ are concave linearly homogenous functions of the price vector $p$, and $\varphi(p)$ is a differentiable homogenous of degree zero function of prices.

We further follow Banks, et al. (1997) by giving specific functional forms to $\ln(a(p))$, $b(p)$ and $\varphi(p)$. In particular,

$$\ln(a(p)) = a_0 + \sum_{i=1}^{n} \alpha_i \ln(p_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln(p_i) \ln(p_j)$$

$$b(p) = \prod_{i=1}^{n} p_i^{\beta_i}$$

$$\varphi(p) = \sum_{i=1}^{n} \varphi_i \ln(p_i).$$

The homogeneity conditions required by theory impose certain restrictions on the parameters of the above functions. Specifically,

$$\sum_{i=1}^{n} \alpha_i = 1$$

$$\sum_{j=1}^{n} \gamma_{ij} = \sum_{i=1}^{n} \gamma_{ij} = 0$$

$$\sum_{i=1}^{n} \beta_i = 0$$

\textsuperscript{16}If $\varphi(p)$ is a constant, independent of prices, then the indirect utility function collapses to the class of functions which includes the Almost Ideal Demand System and the Translog model.
Before welfare analysis can be undertaken, the parameters of this utility function must be estimated. This is done by first deriving the demand equations from the utility function. The parameters of the demand system can then be estimated from observed expenditure data.

The demand system we estimate consists of six goods: electricity, gas, food, clothing, alcohol and a category for all other non-durable non-housing expenditures. With the above specification and using Roy’s identity, the share equations for the \( n \) goods are

\[
w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij}^* \ln(p_j) + \beta_i \ln \left( \frac{m_i}{a(p)} \right) + \frac{\varphi_i}{b(p)} \ln \left( \frac{m_i}{a(p)} \right)^2
\]

(2.3)

where \( \gamma_{ij}^* = \frac{1}{2}(\gamma_{ij} + \gamma_{ji}) \). One of the most appealing characteristics of the QUAIDS model is that the share of expenditure on a particular good is a quadratic function of the logarithm of real total expenditure. Therefore, unlike other popular demand systems, Engel curves are not restricted to be linear. Goods can be necessities for a certain range of expenditure and luxuries for another. An informal assessment of our data for gas and electricity (the goods we are most interested in) is supportive of a linear or quadratic relationship between the expenditure shares and the logarithm of real expenditure (see figure 2.1)\(^{17}\). Therefore, the demand system we estimate has a functional form that appears to be consistent with the data.

The QUAIDS system is also parsimonious. There are few additional parameters to estimate over the Almost Ideal system and this last model is nested within it as a special case. When \( \varphi_i = 0 \), the share equation for good \( i \) is the same as that derived from the Almost Ideal specification.

\(^{17}\)The Engel curves of graph 2.1 are non-parametric adaptive kernel estimates. The dashed lines are the bounds of the 95% confidence intervals.
2.3.2 Two-part tariffs and econometric estimation

Before estimation can proceed, the econometric problems posed by two-part tariffs must be addressed.

With fixed charges the share equation (2.3) is no longer valid. A fixed charge is equivalent to a lump sum reduction in income. Therefore, if the consumer still purchases positive amounts of a good after paying the fixed charges, the expenditure share would be,

\[
\left( \frac{p_i x_i}{m - \sum_{j=1}^{n} a_j} \right) = \alpha_i + \sum_{j=1}^{n} \delta_{ij}^{\ast} \ln(p_j) + \beta_i \ln \left( \frac{m - \sum_{j=1}^{n} a_j}{a(p)} \right) \\
+ \frac{\phi_i}{b(p)} \ln \left[ \frac{m - \sum_{j=1}^{n} a_j}{a(p)} \right]^2
\]

(2.4)

where \( a_j \) is the fixed charge (possibly zero) for good \( j \). After paying the fixed
Figure 2.2: Budget constraint with two-part tariff

charges, behaviour should be the same as before, albeit with a lower level of total expenditure. The prices in the vector $p$ should be the marginal prices for the goods.

Figure (2.2) illustrates a consumer's behaviour for the case where one good is sold using a two-part tariff. The budget constraint with linear prices is BE. With a two-part tariff for good $x$ the budget constraint will be (BCD). Demand behaviour (at interior solutions) will be observationally equivalent to the behaviour of a consumer that has an income of $(m - a)$. That is why the share equations (2.4) are the correct specification.

Complications arise, however, when the consumer stops purchasing the good when faced with a fixed charge. Figure (2.3) illustrates this case. It could be possible that the consumer is better off not consuming good $x$ and saving the fixed charge (at utility level $U_0$) than to consume positive amounts.
Figure 2.3: Corner solution with two-part tariff
of both goods at point $F$ (utility level $U_1$). As explained next, the econometric consequences of these corner solutions may be important.

The specification of the above QUAIDS model does not yet include any heterogeneity in household behaviour. We will introduce two sorts of heterogeneity, observed and unobserved. Observed heterogeneity will be accounted for by socio-demographic variables and ownership of durable goods. We shall denote these variables by the vector $x_h$, where households are indexed by $h$. Unobserved heterogeneity will be designated by the variables $\nu_{ih}$. These variables represent differences in preferences, needs or other random influences which are not captured by the observable variables. The vector of unobserved heterogeneity for all goods is $\nu_h$ and $\nu_h \in \mathbb{R}^n$. We make the following assumptions regarding unobserved heterogeneity,

$$E_h [\nu_h | x, p, m] = 0$$  \hspace{1cm} (2.5)

$$E_h [\nu_h \nu_h' | x, p, m] = \Omega.$$  \hspace{1cm} (2.6)

We further assume that the intercept of each share equation depends on the vector $x_h$ and $\nu_{ih}$,

$$\alpha_{ih} = \alpha_i + g_i(x_h) + \nu_{ih}.$$  

With these assumptions the $a(p)$ function of the QUAIDS indirect utility function will be household specific, $a(p, x_h, \nu_h)$. The share equations are now

$$\tilde{w}_{ih} = \alpha_i + g_i(x_h) + \sum_{j=1}^{n} \delta_{ij} \ln(p_j) + \beta_i \ln \left( \frac{\tilde{m}_h}{a(p)} \right)$$

$$+ \frac{\varphi_i}{b(p)} \ln \left( \frac{\tilde{m}_h}{a(p)} \right)^2 + \nu_{ih}$$  \hspace{1cm} (2.7)

where $\tilde{w}_{ih}$ and $\tilde{m}_h$ are the share off good $i$ and expenditure for household $h$ all net of the fixed charges.
The econometric implications of corner solutions can now be explained. In a particular sample, the variable $\tilde{w}_{ih}$ will be censored if some households are better off not consuming the good and saving the fixed charge\textsuperscript{18}. If zero observations are frequent in the data and no account is taken of this phenomenon, the econometric results will be biased. In particular, if the share equations are estimated by OLS using only those observations that have positive expenditure for all goods, the basic condition for consistency,

$$E_h [v_h | x, p, m] = 0,$$

does not hold\textsuperscript{19}. To show this, notice that faced with a fixed charge (for simplicity only on good $k$), household $h$ will decide to purchase the good if

$$\left( \frac{\ln(m_{kh} - a_k) - \ln(a(p, x_{kh}, \nu_h))}{b(p)} \right)^{-1} + \varphi(p) \geq \left( \frac{\ln(m_{kh}) - \ln(a(p^r, x_{kh}, \nu_h))}{b(p^r)} \right)^{-1} + \varphi(p^r)$$

(2.8)

The left hand side is the utility from the interior solution (point F in figure (2.3)) while the right hand side is the utility from not purchasing. This last utility is calculated as the indirect utility of a consumer faced with a linear price vector which would make consumption zero. That is, $p^r$ is the (linear) reservation price vector. In figure (2.3) this price is given by the slope of the tangent to the utility curve at point B.

The reservation price will depend on the shape of the utility function and is, therefore, a function of the parameters, observable variables and the random unobservables. Equation (2.8) gives, implicitly, the region of the space of the unobservable vector $\nu_h$ for which households will be observed.

\textsuperscript{18}Even with linear prices a household might be observed at a corner solution. A frequent example is non-participation in the labour market. In that case the wage rate is below the reservation wage of an individual, and thus hours worked is zero.

\textsuperscript{19}If, on the other hand, the full sample is used (including zero observations) the results will still be inconsistent.
to be consume positive amounts of gas. Call this subspace $\Theta \subset \mathbb{R}^n$. The complement of $\Theta$, $\overline{\Theta}$, will be the region of the unobservable parameter space where it is optimal not to purchase the good. Therefore, if the share equations are estimated with only the positive observations, the expected value of the residuals will not be zero since only a particular subset of these residuals are observed for positive purchasing households\(^{20}\). Formally,

$$E_h [v_h \mid x, p, m, \bar{w}_{\text{gas}} > 0] = 0.$$  

Econometric techniques for dealing with corner solutions in demand systems have been developed\(^{21}\). However, most of these approaches assume a convex budget constraint and exploit the resulting Kuhn-Tucker conditions of the household’s optimisation problem. The case of a non-convex budget constraint is more difficult, but not insurmountable. An outline of a likelihood approach in the context of our model would be the following. First, make a distributional assumption regarding the vector of unobservables, say

$$\boldsymbol{v} \sim \mathcal{N}(0, \Omega).$$

The likelihood contribution of a household that is observed to consume all goods is,

$$l_h = f(v_h; \Omega)$$

(2.9)

where $f(\cdot)$ is the multivariate normal density. The density 2.9) can be transformed into a function of observables and the parameters of interest using equations (2.7). For those observations not consuming a particular set of goods, the likelihood contribution is

$$l_h = \int_{\overline{\Theta}} f(v_h; \Omega) d\Theta.$$  

(2.10)

\(^{20}\)In the unlikely event that the censoring was random or if the distribution of unobservables was symmetric and the censoring was also symmetric around the mean, then the this would not hold.

\(^{21}\)See Wales and Woodland (1983), Lee and Pitt (1986). See also Hausman (1985) and Pudney (1989) for more general reviews of the econometrics of non-linear budget constraints.
That is, the probability mass of drawing a set of unobservables such that, by condition (2.8), it is optimal not to purchase the particular set of goods. Using (2.9) and (2.10) it is straightforward to form the likelihood function for the whole sample.

In our sample all households have electricity connections, but there is a significant proportion, 22%, without gas connection. Therefore, there are many gas shares that are equal to zero and this is the principal censoring phenomenon encountered in our data. As discussed above, if only positive share observations are used, then regression results will be biased. Intuitively, the problem is that in the sub-sample of observations with positive gas shares there will be an overrepresentation of households that have a strong "taste" for gas consumption. Households with a strong "taste" for gas will be more likely to be connected to the gas network and also spend more on gas.

If non-connection to the gas grid were a voluntary economic decision made by households, and only depended on the censoring rule (2.8), the model could be estimated by the maximum likelihood approach outlined above.

Unfortunately, with our specification it is computationally difficult to obtain the set $\Theta$ from condition (2.8). This set would have to be found numerically for each trial parameter vector in the optimising routine. But more importantly, we believe that the non-consumption of gas may be due to other factors which are not captured by condition (2.8). For example, a particular household may reside in an area without a gas supply grid. In addition, even in areas were there is a supply grid, there is usually a connection charge for a previously unconnected household. We have no information on these charges. Also, the installed appliances in a home may also limit the choice of fuel.\footnote{Until now we have not mentioned the possible endogeneity of the choice of appliances, especially central heating. Ideally we should model the choice of appliances and fuel simultaneously (See Dubin and McFadden, 1984). However, we expect the flexible approach described below to correct for some of the biases of not modelling appliance choice. Most of the variables that we condition on in the gas participation equation are going to affect also the appliance choice: region (proxy for home quality and availability of supply),}
For the above reasons we chose a more flexible estimation strategy than the structural full information approach outlined above. We specify a simple probit equation for the gas participation decision,

\[ c^* = z' \beta_c + \mu \]  \hspace{1cm} (2.11)

where \(z\) is the vector of conditioning variables for the household and we observe positive consumption of gas if \(c^* > 0\). We further assume that \(\mu\) and the \(\nu/s\) are multivariate normally distributed. These assumptions enable us to estimate consistently the share equations for gas and electricity by the well known two step Heckman method (Heckman, 1979).

The approach has the advantage that it allows for other unmodelled-factors, such as fixed costs, cost of connection to the supply grid or cost of energy using appliances to affect the connection decision. The disadvantage of this approach is that efficiency is reduced by not taking into account the cross equation restrictions between the demand equation and the censoring rule if this rule is somehow based on the same preference parameters as the demand function. In addition, the assumption that \(\mu\) is normally distributed might seem heroic.

The distributional assumptions underlying the Heckman approach imply that the expected value of the residuals in the share equations, conditional on gas expenditure being positive, has a well defined form. Namely,

\[ E_h [\nu_i \mid x, p, m, \tilde{w}_{\text{gas}} > 0] = \rho_i \sigma_i \lambda_h \]

where \(\rho_i\) is the correlation coefficient between \(\mu\) and \(\nu_i\), \(\sigma_i\) is the standard deviation of \(\nu_i\) and \(\lambda_h\) is the hazard rate for household \(h\) defined as

\[ \lambda_h = \frac{\phi(z_h \beta_c)}{1 - \Phi(z_h \beta_c)} \]  \hspace{1cm} (2.12)

where \(\phi\) and \(\Phi\) are the density and distribution functions of a standard normal.

Income, tenure type, etc.
If the hazard rate, or a consistent estimate, is appended to each share equation as an additional regressor, then the expected value of the residuals of the share equations among positive expenditure observations will be zero. Thus, consistent parameter estimates can be obtained by estimating the share equations using regression analysis on the subset of the data observed to be consuming positive amounts of gas.

2.3.3 Data

The model was estimated using data from the Family Expenditure Survey from 1985 to 1993. The FES is a yearly random survey of about 7,000 households in the United Kingdom. It contains detailed information on household expenditure, income, ownership of durables, demographic characteristics and other variables. Particularly important for the present study is the information on electricity and gas supply connections to the household's dwelling, gas and electricity expenditure, mode of payment of gas and electricity bills, type of central heating and the holdings of energy using durables.

In the FES there is also information on the expenditure on other domestic fuels such as coal and oil. However, modelling the consumption of these alternative fuels is difficult. They are usually purchased in large quantities and stocked. In the sample we observe a few households with large expenditures that do not reflect their consumption during the fortnight of the survey. Conversely, some households that have zero expenditure may be consuming these fuels but were not observed purchasing during the survey period. In the case where all households consume some of these fuels, this infrequency of purchase problem could be overcome (See Meghir and Robin (1992)). However, in the sample there are many households who are genuine non-consumers of coal and oil and there is no way to distinguish them from those households that consume but are not observed purchasing the product. For this reason, other domestic energy goods besides gas and electricity are
ignored. Moreover, gas and electricity are by far the most important items in domestic fuel expenditure. They accounted for 88% of total household fuel expenditure at the beginning of the sample period and rose to 93% in 1993 (ONS, 1995). Spending on alternative fuels is quite marginal, and their omission, we hope, should not significantly bias the results.

There are several categories of consumers; those paying standard tariffs, either through an automatic bank debit system or through a periodic bill. There is an additional category, those that have a coin or electronic prepayment meter. A large portion of these households have had payment problems in the past and were given a meter as an alternative to disconnection. The firm can recoup the debt, as well as the costs of additional fuel, using these meters. Tariffs in this segment of the market are different from the standard tariff. Sometimes they are customer specific, since meters may be set to generate revenues to pay a household’s outstanding debt. Because of the peculiarities of the prepayment sector of the market, our dataset does not include households in this category for either fuel.

For those consumers paying through a direct debit system, the recorded expenditure does not necessarily coincide with their actual consumption for that quarter. This is because the direct debit arrangement is used to smooth expenditure over the year. The amount debited from the customer’s bank account each month is an estimate of the average yearly consumption rather than actual consumption during the period. Fortunately, the FES also records the information from the last gas and electricity advice. The advice records the true consumption by the household during the period and differs from the actual direct debit payment. The advice figures were used as the relevant expenditure for those households that paid through a direct debit system.

One of the novelties of the dataset is the use of two-part tariff information for gas and electricity instead of aggregate price indices for these goods from the Retail Price Index (RPI). Actual tariffs, disaggregated by region and the
fixed and variable element, were obtained from British Gas and the Electricity Association. For the other four goods, price indices were constructed using indices from the RPI.

Since the recorded energy expenditures during the interview period of the FES usually refers to consumption that occurred in the last quarter, prices for electricity and gas have been lagged three months\(^{23}\). The fixed charge for each of the energy goods was subtracted from the expenditure on these goods. Total expenditure was the sum of these net expenditures plus the expenditure on the four other non-energy goods.

Other conditioning variables were regional dummies, cohort dummies, tenure type, number of rooms in the dwelling, ownership of washing machine, gas or electricity central heating, monthly dummies, average temperature, number of adults in the household, number of retired persons in the household, number of females in the household, number and age of children in the household, a dummy variable that reflects whether part of energy expenditures are paid through the rent, and a trend. There is also a temperature variable included. It is the quarterly average of mean daily air temperature at sea level recorded by the Meteorological Office. There are two temperature series, one for Scotland and the other for England and Wales. Households were assigned one or the other depending on the region in which they reside.

The total sample includes 55,959 observations of which 12,498 did not have a gas supply connection\(^{24}\). For the estimation of the share equations, we follow Banks, et al (1997) and trim observations with shares more than three standard deviations away from the mean.

---

\(^{23}\) There is an issue about the correct way to deal with the irresolvable problem of differently dated expenditures. An alternative specification was to contemporaneous prices for all goods. However, the results from that specification were not very different from that presented above.

\(^{24}\) Almost 100\% of households had electricity connections. The few that did not were dropped from the sample. Those households with zero recorded electricity expenditures were also dropped.
2.3.4 Estimation results

The model is estimated in two steps. First, the parameters \( \beta_c \) of equation (2.11) are estimated. We define a dummy variable taking a value of one if the household has a gas connection and zero otherwise. Using Probit analysis on this variable we obtain estimates of \( \beta_c \).

Table (2.1) shows the results of the Probit analysis. The explanatory variables are regional dummies (Scotland is the omitted case), tenure type (owner, rent free or council housing, with rented accommodation as the omitted case), the gas standing charge, the electricity standing charge, the unit price of electricity, gas, food, clothing and alcohol. All prices have been deflated by the Retail Price Index. Additional variables included the number of rooms occupied, the number and age of children, the number of adults, the number of retired people and the number of females in the household. Cohort dummies for the head of household were also included. Five cohort dummies were defined according to the year of birth of the head of household. Since all households in the sample had electricity connections, the expenditure variable used was real non-durable expenditure net of the fixed charge for electricity.

In many cases the household does not make a voluntary decision with respect to the connection to the gas main. For example, a particular household might live in an area without a gas supply grid. The regional dummy variables serve to control for these geographical non-voluntary disconnections.

Strong assumptions are required to rationalise the gas connection equation. First, the data does not contain information on the date the decision to connect to the gas supply network was taken, only whether at the time of the survey the household was connected or not. Since prices are contemporaneous to the survey date, we have to assume that households had perfect foresight when making their connection decision (Dubin and McFadden (1984) make similar assumptions). Second, there are certain variables
Table 2.1: Probit Estimation of Gas Connection Equation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>s.e.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.6420</td>
<td>0.1863</td>
<td>-3.45</td>
</tr>
<tr>
<td>North</td>
<td>0.4768</td>
<td>0.0305</td>
<td>15.64</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>0.7162</td>
<td>0.0281</td>
<td>25.52</td>
</tr>
<tr>
<td>Northwest</td>
<td>0.9468</td>
<td>0.0287</td>
<td>33.97</td>
</tr>
<tr>
<td>East Midlands</td>
<td>0.5535</td>
<td>0.0302</td>
<td>18.32</td>
</tr>
<tr>
<td>West Midlands</td>
<td>0.6447</td>
<td>0.0287</td>
<td>22.48</td>
</tr>
<tr>
<td>East Anglia</td>
<td>-0.0019</td>
<td>0.0328</td>
<td>-0.06</td>
</tr>
<tr>
<td>London</td>
<td>1.0700</td>
<td>0.0294</td>
<td>36.44</td>
</tr>
<tr>
<td>South West</td>
<td>0.0623</td>
<td>0.0266</td>
<td>2.35</td>
</tr>
<tr>
<td>South East</td>
<td>0.5693</td>
<td>0.0233</td>
<td>24.42</td>
</tr>
<tr>
<td>Wales</td>
<td>0.1992</td>
<td>0.0313</td>
<td>6.37</td>
</tr>
<tr>
<td>Owned</td>
<td>0.4496</td>
<td>0.0245</td>
<td>18.36</td>
</tr>
<tr>
<td>Rent free</td>
<td>-0.2610</td>
<td>0.0462</td>
<td>-5.61</td>
</tr>
<tr>
<td>Council</td>
<td>0.2289</td>
<td>0.0259</td>
<td>8.84</td>
</tr>
<tr>
<td>Rooms occp.</td>
<td>-0.0005</td>
<td>0.0030</td>
<td>-0.10</td>
</tr>
<tr>
<td>Children (age 0-1)</td>
<td>0.0912</td>
<td>0.0195</td>
<td>4.68</td>
</tr>
<tr>
<td>Children (age 2-5)</td>
<td>0.0976</td>
<td>0.0255</td>
<td>3.83</td>
</tr>
<tr>
<td>Children (age 6-10)</td>
<td>0.0709</td>
<td>0.0131</td>
<td>5.39</td>
</tr>
<tr>
<td>Children (age 11-18)</td>
<td>0.0368</td>
<td>0.0155</td>
<td>2.42</td>
</tr>
<tr>
<td>Children (age 17-18)</td>
<td>0.1035</td>
<td>0.0493</td>
<td>2.10</td>
</tr>
<tr>
<td>Adults</td>
<td>-0.0012</td>
<td>0.0121</td>
<td>-0.10</td>
</tr>
<tr>
<td>Retired</td>
<td>-0.0347</td>
<td>0.0294</td>
<td>-1.18</td>
</tr>
<tr>
<td>Females</td>
<td>0.0602</td>
<td>0.0160</td>
<td>3.76</td>
</tr>
<tr>
<td>ln(gas standing charge)</td>
<td>-0.0308</td>
<td>0.0147</td>
<td>-2.09</td>
</tr>
<tr>
<td>ln(elec. standing charge)</td>
<td>0.0145</td>
<td>0.0080</td>
<td>1.89</td>
</tr>
<tr>
<td>ln(Pgas)</td>
<td>-0.3709</td>
<td>0.2971</td>
<td>-1.25</td>
</tr>
<tr>
<td>ln(Pelec)</td>
<td>-0.1301</td>
<td>0.1866</td>
<td>-0.70</td>
</tr>
<tr>
<td>ln(Pfood)</td>
<td>1.9151</td>
<td>0.5566</td>
<td>3.44</td>
</tr>
<tr>
<td>ln(Pclothing)</td>
<td>-1.2347</td>
<td>0.3415</td>
<td>-3.62</td>
</tr>
<tr>
<td>ln(Palcohol)</td>
<td>0.1422</td>
<td>0.3578</td>
<td>0.35</td>
</tr>
<tr>
<td>ln(expenditure)</td>
<td>0.1003</td>
<td>0.0470</td>
<td>2.14</td>
</tr>
<tr>
<td>ln(expenditure)^2</td>
<td>-0.0044</td>
<td>0.0048</td>
<td>-0.92</td>
</tr>
</tbody>
</table>

Observations = 55969  
Likelihood Ratio = 4923.6  
McFadden's pseudo-$R^2$ = 0.08  
Percent correctly predicted = 78.05

Notes:
- Dependent variable is takes a value of one if the household has a gas connection and zero otherwise.
- Expenditure was net of the electricity standing charge.
- The fixed charges, the unit prices, as well as net expenditure, have been deflated by the RPI.
- Results for Cohort variables not reported.
affecting connection, such as a special connection fee, which we have not controlled for. However, we expect the regional dummies to capture some of the effects of these omitted variables.

Most parameters have the expected sign. The regional dummies are for the most part positive and strongly significant, pointing to the fact that in Scotland (the omitted case) the use of gas is not as common as further South. As expected, home ownership increases the probability of having a gas connection. Council tenants also have a higher probability of being connected than households in rented accommodation (the omitted case). Households with children, especially young ones, are also more likely to be connected.

Although not shown in the table, the sign on the cohort dummies show that older cohorts are less likely to have a gas connection.

The coefficients on the price variables are more perplexing. Neither the gas nor electricity unit price seem to affect the connection probability. However, the price of food and clothing seem to have a strong impact. We do not have an explanation for this result, except to reiterate that the prices used may be very poor indicators of the price expectations when the decision to connect was made.

The parameter on the gas standing charge has the expected sign and is significant. At the sample mean of the independent variables, this coefficient implies a marginal reduction in the probability of connection of 0.0087 from a 1% increase in the fixed charge. The coefficient on the electricity standing charge is positive but not significant at the 5% confidence level. It implies a marginal increase of 0.0041 in the probability of connection from a 1% increase in the standing charge.

Higher expenditure households are more likely to have gas connection. In this sense, gas supply (as distinct from consumption) would seem to be a normal good. The insignificance of the coefficient on squared expenditure indicates that the effect does not decline for higher expenditure levels.

With the estimates from the probit equation a hazard rate, $\lambda_h$, is con-
constructed for each household. The hazard rate is added to the share equations as an additional regressor. Then the demand system can be estimated using only those observations which are observed to be consuming gas. The new variable \( \lambda_h \) controls for the fact that preference for gas consumption will be higher within this sub-sample than in the overall sample.

The share equations to estimate are

\[
\tilde{w}_{ih} = \alpha_i + g_i(x_h) + \sum_{j=1}^{n} \delta_{ij} \ln(p_j) + \beta_i \ln\left( \frac{m_{ih}}{a(p)} \right) + \frac{\varphi_i}{b(p)} \ln\left( \frac{m_{ih}}{a(p)} \right)^2 + \tau_i \lambda_h + \nu_{ih} \tag{2.13}
\]

which, except for the \( a(p) \) and \( b(p) \) functions, are linear in the variables if \( g_i(x_h) \) is assumed to be linear also. This suggests an iterative estimation routine. With some starting values for the parameters we construct the price functions. Deflate expenditure and estimate new parameters assuming that \( a(p) \) and \( b(p) \) are constant. Use the estimated parameters to construct new price functions and repeat the process until starting values and estimated values converge.

It is important to note that for the parameter \( \tau_i \) to be identified, there must be an exclusion restriction in the demand system. That is, there must be at least one variable which shifts \( \lambda_h \) but which is not a right hand side variable in the share equations\(^\text{25}\). In our model these exclusion restrictions are provided by the gas and electricity standing charge variables which enter the probit equation but not the share equations\(^\text{26}\). These charges should affect the connection decision but, conditional on being connected to the gas

\(^{25}\) Alternatively, one could assume that the non-linear functional form of \( \lambda_h \) is enough to identify \( \tau \).

\(^{26}\) It must be noted that the electricity standing charge variable is not significant in the Probit equation, but the gas standing charge is. This last variable is then providing the needed identification restriction.
network, should not affect the quantity demanded, except perhaps through an income effect.

Conditional on the price deflators, new parameter estimates were obtained using instrumental variables. All regressors were assumed to be exogenous except for the expenditure variables which were instrumented with income, age of head of household, and squared terms of these last variables.

The results for the gas and electricity equation are presented in table 2.2. The results for the other equations are presented in table (2.7) of the appendix.

To be consistent with consumer demand theory, the estimated demand system must be symmetric. That is, the cross price effects between equations must be equal. Symmetry was imposed on the final instrumental variable estimates using a one step minimum distance estimator. This restriction was rejected by the data, as can be seen from the Amemiya statistic presented in the notes to of table 2.2.

Households in most regions of England and Wales have a tendency to consume more gas than their counterparts in Scotland (the omitted dummy). This is the reverse for electricity, pointing to the predominance of electricity as an energy source in the later region. Household characteristics are an important determinant of energy expenditure, especially the central heating type. In the case of electricity we were able to control for the presence of a washing machine in the household and a fridge/freezer, both of which have the expected positive effect on electricity expenditure. A dummy variable for those households that paid services in their rent or communal charge has the expected sign since part of their energy expenditure is paid indirectly. The temperature variable is lagged one quarter and has the expected negative sign.

The hazard rate, $\lambda_h$, was significantly positive in the gas equation. This is consistent with the hypothesis that people that connect to the gas network have an above average preference for gas consumption.
Table 2.2: Gas and electricity share equations with symmetry imposed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gas share equation</th>
<th>Electricity share equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter t</td>
<td>Parameter t</td>
</tr>
<tr>
<td>Constant</td>
<td>0.1638</td>
<td>0.1956</td>
</tr>
<tr>
<td>North</td>
<td>0.0121</td>
<td>-0.0040</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>0.0168</td>
<td>-0.0045</td>
</tr>
<tr>
<td>North West</td>
<td>0.0186</td>
<td>-0.0027</td>
</tr>
<tr>
<td>East Midlands</td>
<td>0.0108</td>
<td>-0.0006</td>
</tr>
<tr>
<td>West Midlands</td>
<td>0.0130</td>
<td>-0.0033</td>
</tr>
<tr>
<td>East Anglia</td>
<td>-0.0020</td>
<td>-0.0031</td>
</tr>
<tr>
<td>London</td>
<td>0.0175</td>
<td>-0.0038</td>
</tr>
<tr>
<td>South West</td>
<td>-0.0022</td>
<td>-0.0012</td>
</tr>
<tr>
<td>South East</td>
<td>0.0048</td>
<td>-0.0029</td>
</tr>
<tr>
<td>Wales</td>
<td>0.0031</td>
<td>-0.0017</td>
</tr>
<tr>
<td>Rooms</td>
<td>0.0036</td>
<td>0.0017</td>
</tr>
<tr>
<td>Gas CH</td>
<td>0.0194</td>
<td>-0.0078</td>
</tr>
<tr>
<td>Elec CH</td>
<td>-0.0072</td>
<td>0.0178</td>
</tr>
<tr>
<td>Washing Machine</td>
<td>0.0061</td>
<td>0.0033</td>
</tr>
<tr>
<td>Fridge</td>
<td>0.0014</td>
<td>0.0064</td>
</tr>
<tr>
<td>Paid in rent</td>
<td>-0.0019</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Temperature_{t-3}</td>
<td>-0.0016</td>
<td>-0.0002</td>
</tr>
<tr>
<td>Trend</td>
<td>-0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>Owned</td>
<td>0.0115</td>
<td>0.0020</td>
</tr>
<tr>
<td>Rent free</td>
<td>-0.0007</td>
<td>0.0003</td>
</tr>
<tr>
<td>Council</td>
<td>0.0073</td>
<td>0.0011</td>
</tr>
<tr>
<td>( P_{gas, t-3} )</td>
<td>0.0145</td>
<td>-0.0143</td>
</tr>
<tr>
<td>( P_{electricity, t-3} )</td>
<td>-0.0143</td>
<td>0.0052</td>
</tr>
<tr>
<td>( P_{food} )</td>
<td>-0.0343</td>
<td>-0.0264</td>
</tr>
<tr>
<td>( P_{clothing} )</td>
<td>-0.0076</td>
<td>0.0000</td>
</tr>
<tr>
<td>( P_{alcohol} )</td>
<td>0.0342</td>
<td>0.0134</td>
</tr>
<tr>
<td>Expenditure</td>
<td>-0.0526</td>
<td>-0.0537</td>
</tr>
<tr>
<td>Expenditure^2</td>
<td>0.0025</td>
<td>0.0033</td>
</tr>
<tr>
<td>( \lambda_n )</td>
<td>0.0442</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

Observations: 40,431
Average sample share: gas = 0.0446; electricity = 0.0366

Notes:

- Results for cohort, month, number of children, number of adults, female and retired persons in household variables not reported.
- The share of the energy goods as well as total expenditure are net of the weekly fixed charges.
- Symmetry imposed through minimum distance estimator. The Amemiya Statistic for symmetry restriction, \( x^2_{10} \), was 38.14.
- The constant for the price deflator, \( \alpha_0 \), was set to zero.
The parameters on the price variables merit some discussion. The own price coefficients in both equations are positive, and, in the case of electricity, statistically insignificant from zero. A positive or zero coefficient is not inconsistent with negative own price elasticities since the dependent variable is the share of the good in total expenditure. For goods that have inelastic demands, the quantity purchased will decrease with a price increase while the share over total expenditure nonetheless rises.

The cross price terms indicate that electricity and gas are complements, food is a complement to both energy goods and alcohol is a substitute to both energy goods. The complementary between gas and electricity is not at all surprising, given that we are assuming that the ownership of energy using durables is fixed. Most electricity using appliances (such as lighting, kitchen appliances, and televisions) are used in household activities where one would also expect gas using appliances to be used (gas cookers, gas central heating and water heating).

Table 2.3 presents the average compensated, uncompensated and budget elasticities for all equations. It is clear that the demand for both gas and electricity are inelastic with respect to their own price. The budget elasticities also show that these goods are necessities over the range of expenditure present in the sample.

Besides symmetry, consumer demand theory places other restrictions on the estimated demand system. To be consistent with the postulates of utility maximisation, the Slutsky matrix should be negative semidefinite. A necessary condition for negative semidefinite of the Slutsky matrix to hold is that the own-price substitution effects be non-positive. The matrix of compensated elasticities reported in table 2.3 does indeed imply negative own-price effects as all the diagonal elements are negative. Unfortunately, this is not sufficient for a negative semidefinite Slutsky matrix. If the Slutsky matrix is evaluated for each household, only 17,671 of the 40,431 observations meet the integrability condition. However, the aggregate Slutsky matrix is indeed
Table 2.3: Weighted average of individual elasticities

<table>
<thead>
<tr>
<th>Compensated</th>
<th>Gas</th>
<th>Electricity</th>
<th>Food</th>
<th>Clothing</th>
<th>Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gas</strong></td>
<td>-0.29</td>
<td>-0.13</td>
<td>-0.23</td>
<td>-0.33</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Electricity</strong></td>
<td>-0.13</td>
<td>-0.56</td>
<td>-0.10</td>
<td>-0.11</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Food</strong></td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.13</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Clothing</strong></td>
<td>-0.12</td>
<td>-0.04</td>
<td>0.75</td>
<td>-1.41</td>
<td>-0.55</td>
</tr>
<tr>
<td><strong>Alcohol</strong></td>
<td>0.41</td>
<td>0.03</td>
<td>0.42</td>
<td>-0.94</td>
<td>-2.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uncompensated</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gas</strong></td>
<td>-0.31</td>
<td>-0.14</td>
<td>-0.31</td>
<td>-0.34</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Electricity</strong></td>
<td>-0.15</td>
<td>-0.57</td>
<td>-0.22</td>
<td>-0.13</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Food</strong></td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.37</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Clothing</strong></td>
<td>-0.17</td>
<td>-0.08</td>
<td>0.30</td>
<td>-1.55</td>
<td>-0.64</td>
</tr>
<tr>
<td><strong>Alcohol</strong></td>
<td>0.35</td>
<td>-0.02</td>
<td>-0.10</td>
<td>-1.08</td>
<td>-2.73</td>
</tr>
</tbody>
</table>

| Budget          | 0.17  | 0.31        | 0.67 | 1.39     | 1.55    |

Notes:

- Individual elasticities are aggregated using household expenditures over total sample expenditures as weights.
negative definite.

With such a large and heterogeneous dataset it is not surprising that Slutsky negativity is rejected for many households\(^\text{27}\). Flexible functional forms such as those specified above for our QUAIDS system are only a second-order approximation to an arbitrary function around the expansion point. Therefore, it is not surprising that our model behaves considerably well around the average (or aggregate) values of our data, but performs rather poorly for a substantial number of individual observations.

The rejection of Slutsky negativity for some individual observations is troubling for the welfare analysis undertaken below. Consistency in choices, as reflected in Slutsky negativity, should be a basic requirement for inferring welfare effects of price changes. However, the rejections seem to be related to off diagonal elements in the Slutsky matrix that are not related to fuel prices.

### 2.4 Welfare Calculations

In this section we apply the model results to estimate the welfare impacts of tariff rebalancing in the domestic gas market. It is difficult to establish the exact magnitude of the cross subsidy. We use the information contained in Burns, et al (1995). They claim -using information from primary sources-that customer related costs of British Gas are in the order of £65. The current standing charge is only £37, implying that a significant (76%) increase in this price would be necessary to reflect costs. On the other hand, the commodity related costs are estimated to be 39 pence per therm, 12% below the 43.8 pence per therm charged by British Gas. Assuming these numbers reflect, approximately, the magnitude of the cross subsidy, we will analyse the welfare impact of three scenarios. The first one -the baseline scenario- reflects the

---

\(^{27}\)Banks, et al (1996) limit their dataset to a fairly homogenous group of households (married couples living in London or the South East with the husband employed) and obtain much better results in terms of the integrability conditions.
information presented above and corresponds to a total rebalancing of tariffs. The other two scenarios involve a higher and lower decrease in the unit price under the constraint that profits be the same as in the baseline case. The three cases analysed are the following:

i) A 76% increase in the standing charge and a 12% decrease in the unit price.

ii) A 50% increase in the standing charge and a 8% decrease in the unit price.

iii) A 105% increase in the standing charge and a 16% decrease in the unit price.

As a welfare measure we use the compensating variation. It corresponds to the amount of monetary resources that must be given to a household after a price change in order for that household to be able to obtain the same utility level that it enjoyed before the change. Formally, the compensating variation is

\[
CV = E(u^*, p^1, a^1) - E(u^*, p^0, a^0)
\]

where \( u^* \) is the initial utility level, \( p^0 \) and \( a^0 \) are the initial variable price vector and the initial vector of fixed charges, and \( p^1 \) and \( a^1 \) are the new price vectors.

Since the parameter estimates from the demand system are also the parameters of the utility function, we are able to estimate the compensating variation for each household. We do this for observations in the final year of the data set (1993).

The results are presented in table 2.4. The majority of households have a negative compensating variation. This means that income must be taken away from these households in order for them to have the same level of welfare as before the price change. In other words, these households are better off after the tariff rebalancing than before. The negative effects of the rise in
Table 2.4: Welfare Effects of Tariff Rebalancing

<table>
<thead>
<tr>
<th>Scenario:</th>
<th>Number of Households with Change in gas consumption</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pos. cv</td>
<td>neg. cv</td>
</tr>
<tr>
<td>i) Case 1:</td>
<td>76% increase in fzd charge</td>
<td>12% decrease in unit price</td>
</tr>
<tr>
<td>ii) Case 2:</td>
<td>50% increase in fzd charge</td>
<td>8% decrease in unit price</td>
</tr>
<tr>
<td>iii) Case 3:</td>
<td>105% increase in fzd charge</td>
<td>16% decrease in unit price</td>
</tr>
</tbody>
</table>

Notes:

- Calculations were made for the observations in the final year of the data set (1993) and for which the expenditure function was well behaved in the gas price. In each scenario, high income observations which had negative predicted shares were dropped. This explains the slight variation in the total households in each scenario.

- A positive compensating variation implies that the household is worse off after the price change than initially, the opposite is true for a negative compensating variation.
the fixed charge are more than compensated by the savings due to the fall in the variable price.

All scenarios presented above are roughly profit neutral\(^{28}\). Profit neutrality implies that some consumers must lose from the tariff change since someone has to pay for the lower expenditure (and profit margins) of the winners. It can be seen from the second column of table 2.4 that there is a significant group of households that are hurt by the tariff changes.

The issue of non-marginal responses when households face a non-linear budget constraint must be addressed at this point. The results of table 2.4 overestimate the welfare impacts of some households that have a positive compensating variation. This is because some of these households might opt to disconnect from the service instead of bearing the full cost of tariff rebalancing. This, however, does not imply that the results of table 2.4 are too unfavourable to tariff rebalancing. It is true that the compensating variation for those households that decide to disconnect is overestimated. However, the revenues lost from these households must be recovered from those that stay in the market. Therefore, tariffs would have to increase further than what we have considered so far if the price change is to be profit neutral. This would entail additional welfare losses for other households.

Figure 2.4 plots the compensating variations (for case 1) against total expenditure. It is clear from the graph that there is a negative relation between the compensating variation and the initial expenditure level. Higher expenditure households tend to gain from the tariff rebalancing while lower expenditure households are disproportionately the ones that suffer. This is not surprising, higher expenditure is associated with a higher consumption of gas (although the share of expenditure on gas is decreasing with total expenditure). The rise in the fixed charge affects all households by the same

\(^{28}\)If they were not, the tariff changes would imply decreases (or increases) in the average price of gas. The welfare inference would then be a mixture of the effects of tariff rebalancing and the effects of a general price change.
Figure 2.4: Compensating variation and total expenditure (case 1).

amount, while the savings due to the decrease in the unit price is greater for households with high consumption levels.

To analyse in more detail who the winners and losers are, table 2.5 presents the results for stratified sub-samples of the data.

The regressive nature of the price changes -already observed in graph 2.4- can be seen by observing the impact by quintile. On average households in the first quintile, the poorest 20% of the sample, suffer a welfare loss. In general, households that consume low amounts of gas will suffer from the rebalancing of tariffs since the standing charge is a larger fraction of their overall bill. Therefore, households with characteristics that are associated with higher gas expenditure, such as the presence of children or higher total expenditure, will have a lower (and usually negative) compensating variation.
Table 2.5: Average compensating variation for different groups

<table>
<thead>
<tr>
<th>Expenditure Quintals</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retired persons</td>
<td>0.06</td>
<td>-0.12</td>
<td>-0.25</td>
<td>-0.38</td>
<td>-0.54</td>
<td>-0.15</td>
</tr>
<tr>
<td>No retired persons</td>
<td>0.07</td>
<td>-0.09</td>
<td>-0.21</td>
<td>-0.31</td>
<td>-0.52</td>
<td>-0.25</td>
</tr>
<tr>
<td>Age of HH &lt; 25</td>
<td>0.09</td>
<td>-0.02</td>
<td>-0.12</td>
<td>-0.06</td>
<td>-0.28</td>
<td>-0.02</td>
</tr>
<tr>
<td>Age of HH &gt; 24 and &lt; 65</td>
<td>0.06</td>
<td>-0.10</td>
<td>-0.22</td>
<td>-0.32</td>
<td>-0.53</td>
<td>-0.27</td>
</tr>
<tr>
<td>Age of HH &gt; 65</td>
<td>0.06</td>
<td>-0.11</td>
<td>-0.24</td>
<td>-0.41</td>
<td>-0.50</td>
<td>-0.10</td>
</tr>
<tr>
<td>No kids in household</td>
<td>0.07</td>
<td>-0.09</td>
<td>-0.20</td>
<td>-0.31</td>
<td>-0.42</td>
<td>-0.15</td>
</tr>
<tr>
<td>Kids in household</td>
<td>0.03</td>
<td>-0.12</td>
<td>-0.26</td>
<td>-0.34</td>
<td>-0.62</td>
<td>-0.34</td>
</tr>
<tr>
<td>All</td>
<td>0.07</td>
<td>-0.10</td>
<td>-0.22</td>
<td>-0.33</td>
<td>-0.52</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

Notes:

- Calculations were made using the compensating variation of the baseline scenario. Values are weekly compensating variations in 1993 prices.

- Each cell contains the average compensating variation for those households in that group. HH stands for Head of Household.

- A positive compensating variation implies that the household is worse off after the price change than initially, the opposite is true for a negative compensating variation.
The welfare impacts also differ according to the position in the life cycle of the head of household. On average, when the head of the household is middle aged (between 25 and 65 years) the compensating variation is lower than is the case for younger or older heads of households.

The results of table 2.4 show that a significant group of households stand to lose from tariff rebalancing. Furthermore, the poor and old would be one of the most vulnerable groups affected. These results are not particularly new (see Burns, et al (1995) and Hancock and Waddams-Price (1995)). However, with the results of the present study some additional questions can be answered with respect to the welfare impacts of tariff rebalancing.

Figure 2.5: Frequency distribution of compensating variation (case 1).

First, how important are the welfare losses for the households that are negatively affected in comparison to the welfare gains of the others? Figure
2.5 shows the distribution of the compensating variation for the price changes of case 1. The positive compensating variations are small in comparison to the negative compensating variations. This implies that the winners gain more than the losers suffer and thus, potentially, the winners could compensate the latter group and still have a positive net welfare gain.

Another interesting question refers to the combined effect of tariff rebalancing and general price cuts that competition might produce. There is evidence that the introduction of competition in the industrial gas market reduced prices between 10% and 15% (House of Commons (1994)). Some independent suppliers have declared that they will be able to substantially reduce the standing charge for domestic gas customers once they are allowed to supply in a competitive market (House of Commons (1994)). In addition, the regional electricity companies are entering the gas supply market and there may soon be an integrated domestic energy charging system. This could produce significant economies in billing and other customer related costs. Therefore, it is not unrealistic to assume that cost savings could be achieved once competition in the supply of gas intensifies. The crucial question is, how large would these costs savings have to be in order to benefit those consumers that lose from the tariff rebalancing?

Table 2.6 shows the effects of tariff rebalancing (case 1) combined with alternative price reductions. The first three cases assume that competition will reduce customer related costs by 10%, 20% and 30% respectively, and that these cost reductions are passed on to consumers in the form of lower fixed charges. The final case assumes that customer related costs are reduced by 20% and commodity related costs are reduced by 10%. It is clear that, if competition is able to reduce customer related costs (from the current £65) by between 20% and 30%, the vast majority of consumers would benefit even after tariffs are rebalanced.

One last comment regards the short run nature of our estimated demands. The energy demand equations were conditioned on certain durable
Table 2.6: Net welfare effects of tariff rebalancing (case 1) combined with cost reductions

<table>
<thead>
<tr>
<th>Scenario:</th>
<th>Number of Households with Change in gas consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pos. cv</td>
</tr>
<tr>
<td>i) Case 4: 10% decrease in customer related costs</td>
<td>767</td>
</tr>
<tr>
<td>ii) Case 5: 20% decrease in customer related costs</td>
<td>310</td>
</tr>
<tr>
<td>iii) Case 6: 30% decrease in customer related costs</td>
<td>65</td>
</tr>
<tr>
<td>iv) Case 7: 20% decrease in customer related costs, 10% decrease in supply related costs</td>
<td>63</td>
</tr>
</tbody>
</table>

Notes:
- Calculations were made using the tariff rebalancing scenario of case 1. Obviously, the cases presented above are not revenue neutral.
goods. This eliminates the modelling and estimation problems posed by energy consuming durables in the demand for fuels. But the cost is that the demands, and the elasticities calculated from them, are only relevant in the short run. In the long run, households will be able to change their holdings of durables and we have not considered that possibility here. However, since consumers will adapt their behaviour in order to avoid negative impacts, any long-run substitution possibilities in durable ownership for consumers must decrease any negative impact of tariff rebalancing.

2.5 Conclusions

From the results presented in the previous section we may conclude that the rebalancing of tariffs between the fixed charge and the variable charge will have important distributional impacts. Short of avoiding the tariff rebalancing, there are several alternatives to address the distributional consequences of this change. In the first place, as argued by Burns, et al. (1995), the existing welfare system could be used as a compensating instrument. Some form of energy subsidy already exists in the present system. Pensioners, for example, receive benefits that are linked to weather conditions as a way to compensate for higher energy bills. Programs like this could be expanded if negative distributional effects on some groups are to be avoided by tariff restructuring. The advantage of these programs is that they address distributional issues through the welfare system and thus allow prices to be set by efficiency considerations only. We have shown that the compensation required by those who stand to lose from the price changes is small relative to the gains of the winners.

A second alternative are lifeline rates. Since the rebalancing of tariffs will affect negatively those households that consume small amounts of the commodity in question (for whom the standing charge is a higher fraction of the total expenditure), some special tariff could be designed for small
consumers.

More importantly however, if competition increases the efficiency of gas supply, it is possible that almost all households gain even when tariffs are rebalanced. Our results show that reducing customer related costs by 20% to 30% will eliminated most of the negative distributional impacts of tariff rebalancing.
Table 2.7: Food, Clothing and Alcohol share equations with symmetry imposed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Food share</th>
<th>Clothing share</th>
<th>Alcohol share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.5120</td>
<td>0.0497</td>
<td>-0.0732</td>
</tr>
<tr>
<td>North</td>
<td>0.0252</td>
<td>0.0168</td>
<td>-0.0023</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>0.0338</td>
<td>-0.0306</td>
<td>-0.0141</td>
</tr>
<tr>
<td>North West</td>
<td>0.0393</td>
<td>-0.0396</td>
<td>-0.0091</td>
</tr>
<tr>
<td>East Midlands</td>
<td>0.0308</td>
<td>0.0305</td>
<td>-0.0181</td>
</tr>
<tr>
<td>West Midlands</td>
<td>0.0340</td>
<td>-0.0285</td>
<td>-0.0172</td>
</tr>
<tr>
<td>East Anglia</td>
<td>0.0062</td>
<td>-0.0073</td>
<td>-0.0077</td>
</tr>
<tr>
<td>London</td>
<td>0.0074</td>
<td>-0.0331</td>
<td>0.0036</td>
</tr>
<tr>
<td>South West</td>
<td>0.0051</td>
<td>-0.0100</td>
<td>-0.0077</td>
</tr>
<tr>
<td>South East</td>
<td>0.0049</td>
<td>0.0356</td>
<td>0.0027</td>
</tr>
<tr>
<td>Wales</td>
<td>0.0131</td>
<td>-0.0096</td>
<td>-0.0017</td>
</tr>
<tr>
<td>Rooms</td>
<td>0.0012</td>
<td>-0.0037</td>
<td>0.0037</td>
</tr>
<tr>
<td>Gas CH</td>
<td>-0.0009</td>
<td>0.0010</td>
<td>-0.0062</td>
</tr>
<tr>
<td>Elec CH</td>
<td>-0.0072</td>
<td>0.0033</td>
<td>0.0069</td>
</tr>
<tr>
<td>Washing Machine</td>
<td>0.0097</td>
<td>0.0063</td>
<td>0.0135</td>
</tr>
<tr>
<td>Fridge</td>
<td>-0.0042</td>
<td>-0.0071</td>
<td>-0.0030</td>
</tr>
<tr>
<td>Paid in rent</td>
<td>-0.0033</td>
<td>0.0004</td>
<td>-0.0006</td>
</tr>
<tr>
<td>Temperature1-3</td>
<td>0.0005</td>
<td>0.0002</td>
<td>-0.0000</td>
</tr>
<tr>
<td>Trend</td>
<td>-0.0003</td>
<td>0.0003</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Owned</td>
<td>0.0306</td>
<td>-0.0139</td>
<td>-0.0216</td>
</tr>
<tr>
<td>Rent free</td>
<td>-0.0183</td>
<td>0.0191</td>
<td>-0.0055</td>
</tr>
<tr>
<td>Council</td>
<td>0.0300</td>
<td>-0.0075</td>
<td>-0.0112</td>
</tr>
<tr>
<td>$ P_{gas}p-3 $</td>
<td>-0.0343</td>
<td>-0.0078</td>
<td>0.0034</td>
</tr>
<tr>
<td>$ P_{electricity}p-3 $</td>
<td>-0.0266</td>
<td>0.0000</td>
<td>0.0134</td>
</tr>
<tr>
<td>$ P_{food} $</td>
<td>0.1578</td>
<td>0.0029</td>
<td>0.0115</td>
</tr>
<tr>
<td>$ P_{clothing} $</td>
<td>0.0327</td>
<td>-0.0525</td>
<td>0.0066</td>
</tr>
<tr>
<td>$ P_{alcohol} $</td>
<td>0.0279</td>
<td>-0.0665</td>
<td>0.0155</td>
</tr>
<tr>
<td>Expenditure</td>
<td>-0.0481</td>
<td>0.0357</td>
<td>0.0055</td>
</tr>
<tr>
<td>Expenditure$^2$</td>
<td>-0.0068</td>
<td>0.0000</td>
<td>-0.0073</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.0096</td>
<td>0.0025</td>
<td>-0.0105</td>
</tr>
</tbody>
</table>

Observations: 40,631

Notes:

- Results for cohort, month, number of children, number of adults, female and retired persons in household variables not reported.
- The share of the energy goods as well as total expenditure are net of the weekly fixed charges.
Part II

Regulation: Supply issues
Chapter 3

Empirical study of the Norwegian bus industry

3.1 Introduction

The theory of incentives in regulation has become a well developed theoretical field. However, empirical research has not been very extensive. Besides Wolak (1994), Wunsch (1994) and Thomas (1995), we are not aware of other empirical studies that deal with the topic of asymmetric information, incentives and regulation from an econometric perspective\(^1\).

In this chapter we use panel data on the Norwegian bus transport industry to investigate the effects of different incentive contracts and their economic implications. One feature of this data set is that firms not only differ with respect to the regulatory regime they face, but they also produce in a related but unregulated market. We hope to exploit this information to estimate the power of current regulatory contracts. A structural model is developed which includes an adverse selection and a moral hazard variable in the firms’

\(^1\) Another strand in the literature is the use of parameterised models such as Schmalensee (1989), Gasmi, Ivaldi and Laffont (1994) and Gasmi, Laffont and Sharkey (1996). This last article combines an engineering cost model with some calibrated functions to design an optimal mechanism. Mathios and Rogers (1989) study the effects of different regulatory schemes on long distance telephone rates in the United States. However, they used a reduced form econometric model and no attempt was made to model the structural relationship between the regulators and the firms. A review of these, and other papers, is presented in chapter 1.
cost function. Parameters estimated from this model can be used to design an optimal mechanism when, as is usually the case, there is asymmetry of information between the firms and the regulator. We use the results from our estimation to simulate the economic effects of different regulatory schemes, including the tendering of bus routes.

The results have immediate policy implications given that the Norwegian government has the intention of introducing tendering as a regulatory mechanism in the bus transport sector. Welfare calculations using our model show that an auction for a fixed price contract achieves almost the same welfare level as an auction of the optimal contract. Therefore, the welfare gains from implementing an optimal auction may not be worth the added complexities that this option would entail. Another interesting result refers to the expected savings to the government in transfers from implementing a fixed price contract auction. Using our model we estimate these savings to be between 650 and 1050 million NOK (depending on the number of bidders that participate in the auction). The government’s implicit estimate of savings are 700 million NOK, which falls within our estimated range.

The next section reviews some of the theoretical developments of incentive theory in regulation. Then, we discuss some aspects of the Norwegian transport industry, with special emphasis on regulatory institutions. The ensuing section presents the model and estimation results. This is followed by the simulations of the different regulatory regimes. The chapter ends with some conclusions and caveats.

3.2 Incentives in regulation

In this section we briefly review some of the main features of the theory of incentives in regulation which has been developed over the last ten years. A more thorough discussion can be found in chapter 1. The purpose here is to briefly review the topic in order to provide some guidelines for our modelling
approach of the Norwegian bus-transportation industry. The emphasis is on the importance of the power of incentive schemes in regulation.

The new theory of regulation offers a normative framework for analysing regulatory policies subjected to informational constraints. Regulated firms are assumed to have private information about production costs. For instance, bus companies will be better informed about the need for drivers in a transport network. Firms can also take discretionary actions (effort) that affect cost of production. Aspects of these actions—such as bus drivers’ time of rest—are normally hard to observe for the regulator.

From the regulator’s point of view, firms should produce efficiently; in the sense that outputs are produced at lowest possible costs (taking into account firms’ disutility of effort), and, due to the cost of public funds, firms should not earn positive rents. However, when informational constraints of the regulator are taken into account, there appears to be a fundamental trade-off between these two goals. If firms are to break-even, given an allowed rate of return on capital, transfers must be equal to costs. A so-called cost-plus contract guarantees that, no matter what is the (unobservable) efficiency of the firm, it will not earn positive rents. The problem of a cost-plus regime, however, is that firms have no incentives to produce efficiently. Any cost saving from increased effort is passed on to consumers. Therefore, effort exerted is low and the result is a higher cost of production.

On the other hand, firms can be induced to produce efficiently by making them residual claimants to profits. Any cost savings due to higher effort accrues to the firm. Therefore it will increase effort until the marginal cost savings equals the disutility of effort, which happens to be also the socially efficient effort level. Such high-powered incentive contracts, however, generally imply that firms earn excessive rents. The reason is that when the regulator does not know the exact cost of production and the firm must earn

\[^{2}\text{Cost plus contracts are analogous to rate of return regulation, or average cost regulation.}\]
a fair rate of return on capital, the now fixed amount of transfers must be based on a pessimistic estimate of the firms’ (unobserved) efficiency. Firms that happen to be more productive will consequently enjoy positive rents.

In a second best world of asymmetric information (and moral hazard) the optimal contract will trade-off some of the efficiency incentives of high powered incentive schemes for the rent extracting properties of a cost plus regulation. To be more precise on this trade-off and on what the optimal policy looks like, we consider a regulated firm with an aggregate cost-function,

\[ C = C(p, Q, \theta - e) \]  

(3.1)

where \( p \) is the vector of input prices, \( Q \) is the output vector, \( \theta \) is the private information productivity parameter, and \( e \) is the unobservable effort variable. The regulator observes costs, output, and input prices. The productivity parameter is drawn from a cumulative distribution function \( F(\theta) \) on \([\theta, \bar{\theta}]\) with density \( f(\theta) \).

The owners of the firm are assumed to maximise,

\[ U = t - \psi(e), \]  

(3.2)

where \( t \) is the net transfer from the regulator, and \( \psi \) is the disutility of effort function. \( \psi \) is assumed to have the following convenient properties: \( \psi(e) > 0, \psi'(e) > 0, \psi''(e) > 0, \psi'''(e) \geq 0 \).

Assuming a regulator representing the public interest, he will choose output and efficiency so as to maximise expected welfare. Welfare in this case

\(^3\text{A price cap regime, as practised in the United Kingdom, is an example of a high powered incentive scheme. Recent events in that country show how difficult it is for a poorly informed regulator to set caps that do not leave excessive rents to firms. See Armstrong, Cowan and Vickers (1995) for a description of the British regulatory experience.}\)

\(^4\text{See Laffont and Tirole (1993) for a complete description of this problem and chapter 1 for a more detailed derivation than here.}\)

\(^5\text{As in Laffont and Tirole (1993) we make the accounting convention that the government pays the firm’s costs and receives the firm’s revenue. Therefore, } t \text{ the transfer net of cost minus revenues.}\)
is the sum of consumer and producer surplus but we assume that transfers from the regulator to the firm are socially costly. Formally,

\[ W = (S(Q) - R(Q)) + U - (1 + \lambda) (t + C - R(Q)) \]
\[ = S(Q) + \lambda R(Q) - (1 + \lambda) (C + \psi(e)) - \lambda U, \]  

(3.3)

where \( S(Q) \) is total consumer surplus, \( R(Q) \) is revenue, and \( \lambda \) is the cost of public funds.

The regulator's program is to maximise expected (over \( \theta \)) welfare subject to an individual rationality constraint and an incentive compatibility constraint. That is,

\[ \max_{Q, \epsilon} E [S(Q) + \lambda R(Q) - (1 + \lambda) (C + \psi(e)) - \lambda U] \]

subject to

\[ \frac{\partial U}{\partial \theta} = -\psi'(e) \]  
\[ U = t - \psi(e) \geq 0 \ \forall \theta. \]  

(3.4)

(3.5)

The incentive compatibility constraint (3.4) is a consequence of the fact that to implement a social optimal contract it must be designed so that a firm of type \( \theta \) chooses the effort level and output that corresponds to his type. Locally, this will happen if the utility of firms from revealing their true type is not smaller than the utility they gain from pretending to be a different type. The principal must give up rents to firms according to equation (3.4) if they are to reveal their true type.

The individual rationality constraint simply states that the regulator must guarantee non-negative utility for all types of firms\(^6\).

\(^6\)The regulator could also choose to close the market if the firm is very inefficient, but we will not consider that possibility here.
By now familiar steps (see Laffont and Tirole (1993)), the first order conditions for the optimal revelation mechanism are:

\[ S'(Q) + \lambda R'(Q) = (1 + \lambda) \frac{\partial C}{\partial Q} \]  
(3.6)

\[ \psi'(e) = -\frac{\partial C}{\partial e} - \frac{\lambda}{1 + \lambda} \psi''(e) \frac{F(\theta)}{f(\theta)} \]  
(3.7)

From condition (3.7) we can see how the optimal policy balances efficiency considerations and information rents. Given a monotonicity assumption about \( F(\theta)/f(\theta) \), the first order condition implies an effort level which is decreasing in \( \theta \). The reason is that a reduction in effort for a type \( \theta \) depresses the rents given to more efficient types. Effort has this effect on firm's truth telling incentives because the marginal benefit from, locally, announcing a non-true type, \( \psi'(e) \), is increasing with effort. However, for the most efficient type there is no such gain to reducing effort and therefore, for this type, optimal effort is the same as under a first best scenario.

To implement the optimal contract the regulator can offer a non-linear contract designed according to conditions (3.6), (3.7) and (3.2). For example, it could offer a transfer, \( T(C) \), as a function of realised costs and let the firm self select itself by choosing a given cost level. If this transfer function is convex\(^7\), then it can be implemented by a menu of linear contracts (which are tangent to the optimal transfer function). These linear contracts have the general form,

\[ T_i = A_i - b_i(C - C_i) = A'_i - b_iC \]  
(3.8)

where \( C \) is the realised cost, \( A'_i = A_i + b_iC_i \) and \( C_i, b_i \) and \( A_i \) are parameters of the contract. The parameter \( b_i \) defines the power of the contract because it measures how transfers are affected if costs differ from the target cost \( C_i \).

With a carefully designed menu of options the firm that turns out to be of

\[ ^\text{7See Laffont and Tirole (1993) for the relevant assumptions for convexity to hold.} \]
type \( \theta \) will find it optimal to choose a contract such that costs, effort and production are those dictated by conditions (3.6) and (3.7) for that type. The menu of linear contracts has the property that the contract designed for efficient firms has a high \( b_i \) (reaching 1 for the most efficient firm) while the power of the contracts decreases for less efficient firms.

In sum, the optimal contract is such that if the firm happens to be efficient it picks a high powered incentive contract while if it is inefficient it picks a low powered incentive contract. The non-optimality of cost-plus regulation is that it offers a low powered scheme to all types of firms. Likewise, a fixed transfer scheme, such as a price cap, offers a maximal powered scheme to all types of firms.

The empirical problems of designing an optimal contract are vast. The information needed are not easily identified econometrically. For example, the measurement of key parameters, such as the \( \psi \) function, are complicated by the non-observability of effort or efficiency. Therefore, much structure has to be placed on a model in order to estimate these parameters. This will be attempted in section 3.4. However, this paper has also a simpler objective: to investigate whether the predictions from incentive theory help explain differences in productive efficiency among firms. Do high powered regulatory schemes increase efficiency as compared to low powered schemes? If this is so, then incentive theory and modern regulatory theory may be necessary components of production analysis. As Laffont states: "It [New Regulatory Economics] also enables us to model the inefficiencies left so far unmodelled in error terms such as in the econometrics of production frontiers" \(^8\).

---

\(^8\)Laffont (1994), p. 530. The phrase in brackets was added by us.
3.3 Regulatory practice in the Norwegian bus industry

Bus transport is in most countries regulated. Usually each transport company is given exclusive rights to operate a network. A common feature of the industry is the extensive need for transfers from the regulatory authority. In 1989 the share of public subsidies in total revenue for Norwegian transit companies amounted to about 40%. The majority of these companies are privately owned. As in most European countries there is no ex ante competition between firms for network-contracts.\(^9\)

The regulation of the Norwegian transit-industry has undergone several important changes during the last ten years.\(^10\) The main motivation for these changes has been the need to reduce the amount of public funds spent on public transport. In 1983 the old system of ex post balancing of accounts was abolished. It was replaced by ex ante bargaining between the regulator and the companies to fix transfers before production takes place. The responsibility for regulating companies was already delegated to the county-level. Until 1986, however, the central government controlled these counties by targeting county spending on specific activities. This meant that each county received grants from the central government to be used for bus transport only. From the counties point of view, the opportunity cost of public funds in the transport industry was low given that they could not use the resources for other activities. A possible confirmation of this is the fact that the ensuing changes in regulatory practice was initiated by the central government and not by the counties which had the direct responsibility for the

---

\(^9\)An exception is the London Transport System where routes are auctioned for a period of three years. See Kennedy (1994). In Norway, several regulators are now planning to introduce auctions (tendering) for network contracts. Due to EC/EFTA-rules, these auctions will be open to foreign companies.

\(^10\)This section is based on *Nye tilskuddsformer i subsidierede transport* (Report for the Ministry of Transport (1988)) and Andersen (1992).
regulation of the companies. The introduction in 1986 of a block grant system between the central government and the counties was important for the regulation of public transport. This implied that counties were free to move funds between different types of activities. The reform, therefore, increased the opportunity cost of public funds in the transport-industry. At the same time the central government removed all regulatory constraints on the relationship between the counties and the transportation companies.

After this reform, regulatory practice developed differently between counties. Before 1986, all counties practised company-specific bargaining over production level, fares, and costs. From 1986 onwards, some counties introduced a system of standard-costing which removed cost-issues from the company-specific bargaining agenda. In such a system the county and the regulated companies agree upon a set of criteria for calculating costs of operating a bus-network. Given fares and timetables, the standard-cost system gives the level of transfers which is granted by the regulator.

The cost-model is an outcome of bargaining in which certain cost components are linked to the output level, demographic characteristics of the network, stock of capital, and other variables, in a way accepted by all companies in the regulator’s jurisdiction. The important aspect of this scheme is that the same standard costs apply to all companies (within a county) and are not company specific. It is highly plausible then that the standards will be set near the average of the firms’ productivity, thus giving this scheme a certain flavour of yardstick regulation. In any case, a standard cost regime implies a high-powered incentive scheme for the companies: once the cost criteria are fixed, realised costs that deviate from the standardised costs will not influence the level of transfer from the regulator. In light of the new theory of

---

12 On the use of yardstick competition in regulation see Shleifer (1985).
regulation this aspect of the standard-cost regime creates incentives for efficiency, but, depending on the criteria, may leave excessive rents to the transit-companies. If the efficiency standards in the cost-model are set too low, companies will enjoy excessive transfers. This is in fact the main criticism levelled against this regime by Norwegian observers.

Counties which did not adopt the standard-cost system, practice company-specific bargaining about costs and transfers. Compared to companies in the standard-cost regime, these companies are assumed to confront low-powered incentives. Studies of the industry conclude that the regulator’s screening of network-cost is low; the company’s historical costs form the basis for transfers and that incentives for efficiency are low (Ministry of Transport (1988, pp.38-45)). Therefore, this second type of regulatory mechanism resembles a low-power cost-plus regulation.

Given our discussion about the two regulatory mechanisms, we would expect that bus companies regulated by the standard cost approach would be more efficient since they face stronger incentives to increase effort. But there is also another aspect of the Norwegian bus industry that may provide additional information on incentives and efficiency. In each county the regulator controls output by designing the route-network and frequency, fare levels and fare structure in terms of time of day, distance, service type, etc. However, most companies producing local public transport services for the county authority also operate in the unregulated and competitive inter-city transport market. By itself, a competitive market provides maximum incentives to increase effort. Therefore, the unregulated activity will be characterised by high-powered incentives. Due to the close link between the two outputs, the

13This depends, however, on the bargaining process regarding transfers and on the regulator’s ex post reaction to deviations from the bargaining outcome. Putting all the bargaining power on the regulator (as in the new theory of regulation), companies’ incentives for efficiency are determined by the set of contracts (transfer and costs) offered by the regulator. Due to the unobservability of the bargaining process, it is important to note that companies may well face low power incentives in a system with so-called lump-sum transfers in which transfers are determined before production takes place.
aggregate efficiency of bus companies should be expected to be linked to the output mix in these two markets\textsuperscript{14}. In the econometric analysis below we will attempt to exploit this information.

### 3.4 The structural model

In this section we will specify a structural model for the bus transport industry. We will assume that regulation is observationally equivalent to the regulator offering a linear contract of the form (3.8) to the firms. That is, regulation is as if firms are offered,

\[ T_i = A_i - b_i C_r \quad \text{for } i = sc, \text{ in} \]  

where \( C_r \) is the cost attributed to regulated output, \( sc \) stands for standard cost contract and \( in \) stands for individually negotiated contract. Our discussion regarding each of the two regulatory mechanisms implies that \( b_{sc} > b_{in} \); standard cost regulation is high powered relative to individually negotiated contracts.

Utility for a firm is then equal to,

\[
U = \delta T_{sc} + (1 - \delta) T_{in} + R(Q_u) - C_u - \psi(e)
\]

\[
= A_{in} + \delta (A_{sc} - A_{in}) - ((1 - \delta) b_{in} + \delta b_{sc}) C_r + R(Q_u)
\]

\[
- C_u - \psi(e)
\]  

(3.10)

where \( \delta \) takes the value of one if the firm is regulated by the standard cost approach, zero otherwise, \( R(Q_u) \) is the revenue in the unregulated market

\textsuperscript{14}For the regulator of local public transport this introduces additional aspects to the regulatory problem. First, the multi-output character of the production-activity opens the difficult problem of monitoring companies' cost-allocation between activities. Besides the more fundamental problem of dividing total costs on outputs when the activities are not completely separable, there will be a problem of accounting manipulations. Inputs such as labour, capital, fuel, etc. can be incorrectly imputed to the regulated activity for which incentives are lower. Second, the firms could substitute effort between activities. The issue of effort substitution will be discussed further below.
and $C_u$ is the cost attributed to the unregulated output, which has to be $C - C_r$.

One crucial assumption that we make regards the breakdown of costs. In general, the ratio of costs of regulated activities to unregulated activities could depend on many factors, both technological and behavioural. An example of the latter are accounting manipulations to transfer costs from the unregulated sector to the regulated sector. Firms will have an incentive to do so if the transfer is sensitive to the reported costs in the regulated sector, as will happen with low powered incentive schemes. Another example is effort substitution. If effort can be directed to a specific sector, then, for a given overall level of effort, the firm will try to divert most of it to the activities with the highest power scheme. In the context of the bus industry this might be the case if the most efficient drivers, buses and mechanics are assigned to unregulated activities. The firm will be able to keep all the cost savings in that sector as extra profits, while some of the extra costs induced by this substitution in the regulated sector will be recouped in the form of higher transfers.

Therefore, a suitable rule for disaggregating costs can be very complex. However, there is one piece of information which we can be sure about. If unregulated output is zero then regulated costs are the firm's total costs, and if regulated output is zero then regulated costs are zero. Therefore, we could postulate the general relationship as

$$\frac{C_r}{C} = \left( \frac{Q_r}{Q} \right)^{\alpha}.$$  

where $\alpha$ could depend on many factors, including the output levels and the composition of effort. Rather than modelling $\alpha$, below we will estimate a constant value for this variable using a cross section for a particular year where disaggregated cost figures are available. There are several reasons for taking this restrictive approach. First, we lack suitable data for modelling and estimating this parameter. It would be very difficult to identify $\alpha$, as well
as all the other parameters of the model, by using only the aggregate costs observed in the panel. Second, modelling this parameter would complicate an already non-linear model and at this stage we prefer to maintain the model as simple as possible. In any case, as will be discussed below, sensitivity analysis was performed by estimating the model for different values of $\alpha$ without any noticeable changes in the results.

Using the cost breakdown rule, the firm's utility can be expressed as,

$$ U = (A_{in} + \delta(A_{sc} - A_{in}) + R(Q_u)) $$

$$ - (I_u + ((1 - \delta)b_{in} + \delta b_{sc})I_r) C - \psi(e) $$

(3.11)

where,

$$ I_r = \left( \frac{Q_r}{Q} \right)^\alpha $$

$$ I_u = 1 - \left( \frac{Q_r}{Q} \right)^\alpha. $$

Total operational costs are given by,

$$ C = W_dL_d + W_aL_a + P_{fu}fu $$

(3.12)

where $L_d$, $L_a$, and $fu$, are the driver labour, administrative labour and fuel inputs respectively, and $W_d, W_a, P_{fu}$ are the prices of the variable inputs. A well behaved production technology is also assumed,

$$ g(K, L_d, L_a, fu, \theta, \epsilon) \geq Q $$

where $K$ is the capital stock and $Q$ is the output vector. Conditional on a level of effort and output, and assuming $(I_u + ((1 - \delta)b_{in} + \delta b_{sc})I_r > 0$, maximising utility 3.11 is equivalent to minimising costs subject to producing
the given level of output. That is, utility -conditional on effort and output-is maximised by,

\[
\min_{L_d, L_a, f_u} \quad C = W_d L_d + W_a L_a + P_{fu} f_u
\]  

(3.13)

subject to:

\[
g(K, L_d, L_a, f_u, \theta, e) \geq Q.
\]

Therefore standard duality results can be applied and profits can be expressed by the following equation,

\[
U = (A_{in} + \delta (A_{sc} - A_{in}) + R(Q_u))
\]

\[
- (I_u + ((1 - \delta) b_{in} + \delta b_{sc}) I_r) \bar{C} - \psi(e)
\]

(3.14)

where \(\bar{C}\) is a dual cost function conditional on the level of effort and output and is given by,

\[
\bar{C} = C(W_d, W_a, P_{fu}, K, Q, \theta, e).
\]

(3.15)

Maximising (3.14) with respect to effort yields the first order condition that determines optimal effort for each company \(^{15}\),

\[
(I_u + ((1 - \delta) b_{in} + \delta b_{sc}) I_r) \frac{\partial \bar{C}}{\partial e} = \psi'(e)
\]

(3.16)

Equation (3.16) is revealing. If the incentive contracts are very low pow­
ered (\(b's\ close to zero) then optimal effort will only be a fraction of the socially optimal effort level. This fraction would depend on how important unregulated activities are compared to regulated activities. On the other hand, if the power of the contracts are close to one then the expression in

\(^{15}\)We are making the implicit assumption that output, and thus revenue, in the unreg­ulated sector does not depend on effort or the unobserved efficiency parameter.
parenthesis will be close to one and effort will be close to the socially optimal level.

Equations (3.16) and (3.15) form the basis of the structural model. What we want to exploit is the fact that higher activity in the unregulated sector may help to identify the model, since by equation (3.16) effort will be correlated with the proportion of output in the unregulated sector. The intuition is that firms with higher activity in the unregulated sector have a higher incentive to increase effort.

In the rest of this section we will derive an analytical solution to equations (3.16) and (3.15) by making strong functional assumptions. However, it should be noted that in principle a more general approach could be pursued at an extra computational cost.

A specific assumption will be made regarding the way $\theta$ and $e$ enter the cost function. It will be assumed that the production function depends on $L_d^*$ rather than $L_d$, where,

$$L_d^* = \frac{L_d}{\exp(\theta-e)}.$$  \hspace{1cm} (3.17)

$L_d^*$ can be interpreted as efficiency units of driver labour which varies inversely with $\theta$ and positively with the effort level of managers. Only labour was chosen as the variable whose productivity is affected by the adverse selection parameter and the effort variable. In bus-transit, drivers are the most important input in terms of costs. Firms' efficiency-discretion through the choice of effort is expected to be linked to the utilisation of these drivers. Moreover, the relationship between the network structure and the need for driver-hours is reported as one of the most important sources of the regulator's asymmetry of information. The problem is to determine the number of

\footnote{For some production functions, such as the Cobb-Douglas, a simple change of scale of the adverse selection and effort variable will allow these variables to be related to the productivity of any input. Therefore, the choice of input whose productivity is affected is to a certain extent arbitrary.}
hours a driver can spend driving, taking into account the need for using the same driver on different routes, drivers’ waiting-time, check-in and out of buses and time of rest.

If equation (3.17) is substituted into program (3.13) we have an equivalent program,

$$\min_{L^*_d, L_a, f_u} C = W_d \exp(^{(\theta - \epsilon)}L^*_d + W_a L_a + P_{fu} f_u \quad (3.18)$$

subject to:

$$g(K, L^*_d, L_a, f_u) \geq Q.$$  

In this program managers choose the amount of driver labour efficiency units, $L^*_d$. The dual cost function that results from program (3.18) is,

$$C = C(W^*_d, W_a, P_{fu}, K, Q) \quad (3.19)$$

where

$$W^*_d = W_d \exp(\theta - \epsilon).$$

The advantage of equation (3.19) is that now any flexible functional form can be chosen and adapted to the case of asymmetric information and moral hazard. To solve equations (3.16) and (3.15) analytically, however, we will adopt a rather restrictive form, a Cobb-Douglas specification $^{17}$,

$$C = \beta_0 \exp \eta \beta_d W^*_d W_a^{\beta_a} P_{fu}^{\beta_{fu}} K^{\beta_k} Q^{\beta_Q} \quad (3.20)$$

$$= \beta_0 \exp \eta \exp^{\beta_d(\theta - \epsilon)} W^*_d W_a^{\beta_a} P_{fu}^{\beta_{fu}} K^{\beta_k} Q^{\beta_Q} \quad (3.21)$$

$^{17}$In this specification we have to assume that the two outputs can be aggregated. Also, by aggregating outputs we have assumed away the problem of effort substitution.
where we have included a firm effect, $\eta$, reflecting unobservable (to the econometrician) variables that affect costs but are specific to each firm.

With these assumption the effect of an increase in effort is proportional to the cost level,

$$\frac{\partial C}{\partial e} = -\beta \eta C.$$  \hfill (3.22)

We will further assume that

$$\psi(e) = e^{\tau e} - 1 \text{ and } \tau > 0.$$  

With these functional specifications we can solve equations (3.16) and (3.15) to obtain the reduced form cost function,

$$\ln C = \beta + \beta_c \gamma \ln W_d + \beta_a \gamma \ln W_a + \beta_{fu} \gamma \ln P_{fu} + \beta_k \\gamma \ln K + \beta_Q \\gamma \ln Q + (\gamma - 1) \ln (I_u + (1 - \delta) b_{in} + \delta b_{sc}) I_r) + \beta_d \gamma \theta + \gamma \eta \quad \text{(3.23)}$$

where $\gamma = \frac{\tau}{\tau + \beta_c}$. Since $\gamma$ is smaller than one, the parameters of a traditionally estimated cost function will be biased down. This occurs because the impact on costs of a rise in input prices, or output level, will be somewhat mitigated by a rise in effort, since the benefit of increasing effort is proportional to costs. Large firms that produce large amounts of output will have an incentive to increase effort more than small firms. This gives an interesting reason why economies of scale may be overestimated in traditional studies of regulated industries\textsuperscript{18}.

As $\tau$ increases, effort becomes very costly and it's level would tend to be close to zero. In that case, $\gamma$ approaches one and the cost function (3.23)

\textsuperscript{18}Feinstein and Wolak (1991) give another reason why economies of scale may be overestimated in regulated industries. In their model the bias in estimated parameters is induced by the regulator who offers the companies an optimal contract.
collapses to a traditional Cobb-Douglas dual cost function. If the cost of effort is very large then moral hazard is not a relevant phenomenon.

There is still the issue of identification. If the term with the output ratios was linear then the model would not be identified\(^\text{19}\). Therefore, the non-linearity introduced by the logarithmic transformation is crucial to the identification of the model. This is clearly a very weak identifying assumption. In addition, if \(b_{in}\) and \(b_{sc}\) are both close to one (which in our case would be unlikely) the whole term is zero implying that \((\gamma - 1)\) is not identified. To help identify the model, we impose the homogeneity of degree one condition on the cost function before estimation. This guarantees that \((\gamma - 1)\) is identified.

### 3.4.1 Data

The analysis is performed on a balanced panel of 88 Norwegian bus companies over the years 1987-1991. The data are reported annually by the bus companies to the Central Bureau of Statistics. In certain parts of Norway there are large numbers of very small bus companies, often with only one or two employees. As the data-quality of these companies can be questionable, they are excluded from the data-set. Furthermore, companies with incomplete reports in any of the five years are also excluded from the data-set.

Of the 440 observations, 198 were for a company/year regulated by the standard cost approach. There are numerous transitions in the data set as the number of counties adopting the standard cost approach increased.

The number of vehicle kilometres in regulated transport and in unregulated transport are the two outputs used in this study. Labour inputs are measured by total number of hours worked by the drivers and by administration. Wages for these two inputs are derived from reported labour-costs and hours. The fuel-price is calculated from companies reported fuel-costs

\(^{19}\)To see this notice that the constant, \(I_{n}, (1-\delta)I_{r},\) and \(\delta I_{r}\) are not linearly independent.
and quantity used. Capital is measured by the total number of seats of the rolling stock. Summary statistics for the variables are given in table A.

Costs are defined as the sum of driver labour costs, administrative labour costs and fuel costs. Workshop labour, as well as parts and materials, were excluded because of the close connection of these costs to capital expenditure and also the lack of a price variable for parts and materials. Therefore, the function we estimate can be considered a short-run operating cost function.

### 3.4.2 Stochastic specification

There are two variables in the reduced form cost equation (3.23) that are unobservable from the econometrician’s point of view: the firm specific effect \( \eta \) and the adverse selection parameter \( \theta \). We will assume that the efficiency parameters are iid for all firms and years. This means that there is no persistence in the firm’s adverse selection parameter from year to year. Persistence will enter through the individual effect \( \eta \) which we will assume are iid draws for each firm but constant across years. Furthermore, the individual effects are assumed to be uncorrelated with the \( \theta \)'s.

The above assumptions are equivalent to a random effects panel data model. Consistency of parameter estimates require that the individual effects be uncorrelated with all regressors. A more robust alternative would be to estimate the model using a within groups estimator for correlated fixed effects model. However, with such a short panel as ours this requires the alternative assumption that the regressors be strictly exogenous (see Hsiao (1986)). More problematic from a practical point of view, however, is the fact that the fixed effects estimator did not converge. The difficulty would seem to lie in the non-linear term \( ln (L_u + ((1 - \delta)b_{in} + \delta b_{ae})L_r) \). The fixed effect model takes differences from average values of the variables. This makes the estimation of the parameters in the interior of this logarithm term highly unstable. This numerical problem would also seem to apply to estimations based on a first
Figure 3.1: Cost ratio vs Output ratio for 1991 cross section (solid line above 45 degree line is predicted cost ratio from OLS regression).

difference version of equation (3.23) as in the GMM methods for panel data proposed by Arellano and Bond (1991).

3.4.3 Estimation and Results

The first task is to estimate $\alpha$ for the cost breakdown rule. Figure 3.1 shows the cost ratio to output ratio for 1991. For that year, information is available on the cost breakdown for each firm\(^{20}\). The estimate of $\alpha$ from an OLS regression of the logarithm of the cost ratio on the logarithm of the output ratio was .678. The solid line above the diagonal is the predicted cost ratio from this regression. As can be seen from the graph, the proportion of regulated costs to total costs is more than proportional to the output ratio, indicating that some cost shifting between sectors may be occurring\(^{21}\).

With the above estimate of alpha we constructed the $I_u$ and the $I_r$ vari-

\(^{20}\)The cross section contains more firms than the balanced panel used to estimate the cost function.

\(^{21}\)An alternative explanation is that returns to scale are different between the two sectors.
ables for each observation and used these to estimate the cost equation (3.23). Assuming that the fixed effects and the efficiency variable have a normal distribution, the cost equation was estimated as a random effects model using maximum likelihood. Results are presented in table 3.1.

The point estimate of $\gamma$ derived from table 3.1 is 0.5539. Using this value to scale the other coefficients we obtain that $\beta_d = 0.71$, $\beta_a = 0.19$ and $\beta_{fa} = .10$, which are close to the sample average input shares shown in the appendix. Also, once the coefficient on output is scaled by $\gamma$ there appears to be diseconomies of scale. The estimate of $\tau$ is .8852.

In order to see how sensitive the results are to the alpha parameter, the model was estimated using an alpha 10% higher and 10% lower than the base case of .678 used above. The results are shown in the tables in the appendix. Estimates are not very sensitive to variations in $\alpha$.

Contrary to what theory would predict, the coefficient on capital is positive. We rationalise this outcome by arguing that our capital variable is proxying the urban density and the load factor faced by the different companies. We have no information to control for network characteristics that affect costs, such as congestion or peak-hour bus demand. These variables will probably be correlated with the number of seats of the rolling stock. For example, observing a company with many seats may be an indication of high peak demand faced by that company. Congestion may imply that to serve a given level of demand more buses are required to meet timetables. Since congestion, or a low average to peak demand, imply higher costs, then we would expect the parameter on capital to be biased up in our estimation\(^2\).

The estimate for the power of the individually negotiated contract is close to 0.56. This means that this type of regulation is not as high-powered as a pure transfer, but it seems to give firms more incentives to increase effort than a pure cost-plus regulation. As expected, standard cost regulation is more

---

\(^2\)Correlation between capital and the individual effect could also be biasing this coefficient.

161
Table 3.1: Structural cost function estimation results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>s.e.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.0493</td>
<td>0.5236</td>
<td>11.55</td>
</tr>
<tr>
<td>ln Capital</td>
<td>0.1867</td>
<td>0.0256</td>
<td>7.30</td>
</tr>
<tr>
<td>lnQ</td>
<td>0.8234</td>
<td>0.1396</td>
<td>5.90</td>
</tr>
<tr>
<td>lnQ²</td>
<td>0.0040</td>
<td>0.0203</td>
<td>0.20</td>
</tr>
<tr>
<td>ln Wd</td>
<td>0.3948</td>
<td>0.0549</td>
<td>7.20</td>
</tr>
<tr>
<td>ln Wa</td>
<td>0.1032</td>
<td>0.0192</td>
<td>5.37</td>
</tr>
<tr>
<td>γ-1</td>
<td>-0.4461</td>
<td>0.0564</td>
<td>7.91</td>
</tr>
<tr>
<td>b₁ in</td>
<td>0.5583</td>
<td>0.1076</td>
<td>5.19</td>
</tr>
<tr>
<td>b₁ sc</td>
<td>0.6157</td>
<td>0.1147</td>
<td>5.37</td>
</tr>
</tbody>
</table>

Firms = 88  
length of panel = 5  
Mean log-likelihood = 2.448

NOTES:

- Dependent variable is ln C.
- Homogeneity of the cost function has been imposed.
- Output is the simple sum of output in the regulated and unregulated market.
powerful than the alternative mode of regulation, reaching .62. However, the difference is not large nor is it statistically significant.

An implicit assumption in the above estimation is that companies face an exogenous regulatory regime. That is, we assume that companies have no choice as to the regulatory regime they face. However, it is plausible that the regime chosen by a county may itself be subject to bargaining between the regulator and the companies. If this is the case, the efficient firms in a county will pressure the regulator to adopt the standard cost scheme, since under this mechanism efficient firms can earn more rent. This would introduce some correlation between the efficiency parameter, $\theta$, and the last regressor, $\ln (I_u + ((1 - \delta)b_{in} + \delta b_{sc})L_r)$. There is one argument that justifies the assumption that the regulatory regime is not correlated with efficiency. Since under a standard cost approach, transfers will be based on some kind of average of the companies costs, in each county there will be just as many companies opposed to a standard cost regime as those that favour it. Therefore, we do not consider the assumption of regime exogeneity as very restrictive.

3.5 Welfare calculations

The structural model presented in the last section identifies all the relevant information to calculate the optimal regulatory contract. The distribution of $\theta$ can be estimated from the sample residuals. Also, the $\psi$ function is identified, although this was done by assuming a one-parameter family of functions. The only other information needed is the cost of public funds, $\lambda$, which can be borrowed from other studies that try to measure this variable.

---

23Note that we have assumed that the efficiency parameters are not time persistent. Therefore efficient companies in one year might be inefficient the next and they would not have a strong incentive to pressure the regulator for a particular regime. Our model assumptions are then inconsistent with the story behind an endogenous regulatory regime.

24This must be qualified, however. It may be that some companies have more political power than others. It also assumes that companies know their efficiency ranking relative to their rivals.
directly. In this paper we used three values for the cost of public funds, 0.05, 0.3 and 0.5. Our benchmark will be 0.3, which is the lower bound given by the estimates of Pedersen (1994).

We use the estimates from the model to calculate the expected costs and welfare effects of different regulatory regimes. In particular we will compare the properties of the current regimes against an optimal contract, the auction of the optimal contract and a Vickrey type auction for a maximal power incentive scheme. All of these calculations are performed at the average of our dataset for all exogenous variables.\(^{25}\)

Before presenting the results the issue of what constitutes the optimal contract must be addressed. Since there are two sector, the regulated and the unregulated, the results for stylised one sector model presented in section 3.2 are not directly applicable. However, it is easy to show that if the regulator maximises expected welfare from both activities, then the first order conditions for the optimal contract are exactly the same as equations (3.6) and (3.7) provided the cost function is interpreted to be the aggregate costs from both unregulated and regulated activities. We assume that the optimal contract is the contract that maximises welfare over both markets and therefore the optimal contact is characterised by equation (3.7).

From the residuals the distribution of \(\theta\) can be estimated. This approach is similar to the idea proposed by Wunsch (1994) of using the confidence intervals around predicted costs as an estimate of \(F(\theta)\). Ideally, this could be done non-parametrically. However, due to the low number of observations in our data set we have opted to fit a parametric distribution. Figure 3.2 shows the estimated thetas from the model. Except for one outlier the distribution seems symmetric and bell shaped. The outlier was dropped and a normal distribution was fitted. Figure 3.3 shows the estimated thetas (without the outlier) and the fitted normal distribution. The empirical maximum and

\(^{25}\)When calculating the optimal contract we assume that a firm is willing to participate in the regulated sector if its aggregate utility (given by equation (3.11) exceeds zero.
minimum (without the outlier) were used as the bounds on the value of theta.

Using the estimated distribution of $\theta$ and the estimated parameters we calculated the expected costs and welfare under different schemes for the average firm in the data set. We assume both regulated and unregulated output and revenue are constant. Therefore welfare costs are the last two terms of equation (3.3). Table 3.2 presents the results for different values of $\lambda$, the cost of public funds. For each regulatory regime we calculated the expected cost, the expected gross transfers and expected welfare using the effort induced by each regime (solving equation (3.16) for the existing contracts and equation (3.7) for the optimal contract). Effort levels for the optimal regime as well as the welfare cost for all types of contracts will depend on the cost of public funds. Expected costs are the production costs (for a fixed level of output) averaged over the possible $\theta$'s. Expected transfers are production costs plus the cost of effort and rents. This is the total gross transfer that the regulator would have to give to the firm if it wants to
implement each contract\textsuperscript{26}.

The first thing to note is that the standard cost regime lowers costs by about 5\% compared to the individually negotiated contract. However, gross transfers, which include rents, the cost of effort as well as the production cost, is only 2.2\% lower under the standard cost regime. When $\lambda = 0.3$ the welfare cost is lower under the standard cost regime by about 2.7\%. As the cost of public funds increases to 0.5 the welfare advantage of the standard cost regime decreases marginally. Likewise, when rents left to the firm are not costly ($\lambda = 0.05$) then the welfare advantage of a regime that induces more effort is marginally higher. Given that the cost of public funds is very difficult to measure it is reassuring that the results are not very sensitive to this parameter.

Expected costs under the optimal contracts are significantly lower than

\textsuperscript{26}It is not equal to the actual monetary transfer from the regulator for two reasons. First, revenues must be subtracted. Second, an allowance for the return to the capital stock must be included. Therefore, these figures can not be compared to the actual subsidies received by the Norwegian bus industry in a given year.
Table 3.2: Results for different contracts and different costs of public funds at sample average

<table>
<thead>
<tr>
<th>Contract</th>
<th>Expected Values (in 1991 NOKs)</th>
<th>Cost</th>
<th>%</th>
<th>Transfer</th>
<th>%</th>
<th>Welfare Cost</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda = 0.3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind. Negotiated</td>
<td>15,695,273</td>
<td>100.0</td>
<td>20,805,337</td>
<td>100.0</td>
<td>25,608,339</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Std. Cost</td>
<td>14,893,702</td>
<td>94.9</td>
<td>20,347,474</td>
<td>97.8</td>
<td>24,916,356</td>
<td>97.3</td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>9,352,345</td>
<td>59.6</td>
<td>18,578,026</td>
<td>89.3</td>
<td>21,948,852</td>
<td>85.7</td>
<td></td>
</tr>
<tr>
<td>( \lambda = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind. Negotiated</td>
<td>15,695,273</td>
<td>100.0</td>
<td>20,805,337</td>
<td>100.0</td>
<td>20,407,005</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Std. Cost</td>
<td>14,893,702</td>
<td>94.9</td>
<td>20,347,474</td>
<td>97.8</td>
<td>19,829,487</td>
<td>97.2</td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>9,113,309</td>
<td>58.1</td>
<td>18,905,218</td>
<td>90.9</td>
<td>17,274,811</td>
<td>84.7</td>
<td></td>
</tr>
<tr>
<td>( \lambda = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind. Negotiated</td>
<td>15,695,273</td>
<td>100.0</td>
<td>20,805,337</td>
<td>100.0</td>
<td>29,769,406</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Std. Cost</td>
<td>14,893,702</td>
<td>94.9</td>
<td>20,347,474</td>
<td>97.8</td>
<td>28,985,850</td>
<td>97.4</td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>9,474,168</td>
<td>60.4</td>
<td>18,497,924</td>
<td>88.9</td>
<td>25,655,488</td>
<td>86.2</td>
<td></td>
</tr>
</tbody>
</table>

**NOTES:**

- Values are calculated for the average firm in the sample.
- Values are welfare costs, therefore a lower entry implies a higher welfare.
- The percentage columns measures each variable as a percentage of the individual negotiated contract result for each \( \lambda \).
under the current regimes, reaching 58.1% of the individual contract cost when $\lambda = 0.05$. As the cost of public funds rises effort is reduced to lower the rents left to the firm and expected costs are slightly higher.

Figure 3.4 and 3.5 presents the effort and cost level exerted under the different regimes for $\lambda = 0.3$. As expected, the optimal contract sacrifices some efficiency gains for inefficient firms so that informational rents to higher types can be lowered. The optimal contract induces the more efficient firms to exert more effort than under either of the current regulatory regimes. We can see from the graph of costs that these are monotonically increasing in types which guarantees that the second order condition for the optimal contract is satisfied\(^{27}\).

From a policy perspective the important magnitude is the decrease in transfers that result from offering the firms an optimal contract. The results from table 3.2 indicate that, the government could reduce transfers between

\(^{27}\)Costs are monotonically increasing in $\theta$ for the three different values of $\lambda$. Therefore the optimal contracts are feasible.
9.1% to 12.5% (depending on \( \lambda \)) of the total gross transfer under the individually negotiated contract. To predict the effect on current government subsidies it is more convenient to express the transfer savings as a percentage of average operational costs. The savings would then represent between 12.1% and 14.7% of the average production cost under the individually negotiated contract and between 9.7% and 12.4% under the standard cost scheme. The lower bound in the previous figures are for \( \lambda = 0.05 \) and the upper bound for \( \lambda = 0.5 \).

The optimal regulatory contract still leaves informational rents. These rents could be reduced considerably under an auction. Laffont and Tirole (1987) show that the optimal auction would still require the same effort and cost levels from each type as in the optimal contract (the separability principle). However, rents as well as expected costs would be reduced.

The optimal auction could be implemented as a direct mechanism by having firms declare their type and the authorities choosing the lowest type as the winner. This winner would then have to exert the effort allocated to
him by the optimal menu and receive rents equal to the difference between his type and the second lowest type's rents under the optimal contract. Truth telling is a dominant strategy in this auction.

We use the result in section 7.5.1 in Laffont and Tirole (1993) to estimate the effects of this direct auction. These values depend on the number of bidders. It can be seen from table 3.3 that as the number of bidders gets very large, informational rents disappear and the first best (symmetric information) result can be obtained. Even with a few number of bidders, the reduction in informational rent is considerable.

For example, with just two bidders expected informational rents are reduced by more than half. For $\lambda = 0.3$, transfers are reduced to between 89% (when $n = 2$) and 77% (when $n = 100$) of the transfers under an optimal contract. Welfare costs are between 95% and 85% of the costs of the optimal contract. The estimates indicate that, if the optimal contract were to be auctioned, the savings to the regulatory authority - in terms of lower transfers - could be dramatically increased above the 7.7% to 12.5% of operational costs estimated above for the case of an optimal contract.

Due to their complexity, perhaps optimal incentive contracts will never be implemented in the Norwegian transport industry, nor for that matter the auction of this contract. However, the regulatory authorities in Norway are currently considering the introduction of tendering for the bus transport routes. It is interesting to see what our model would predict if such tendering is implemented.

We make the assumption that, instead of the optimal contract, tendering would probably involve awarding the route to the company that makes the lowest bid for a fixed transfer. This would be equivalent to auctioning a fixed price contract.

Table 3.4 shows the results of a second-price Vickrey auction of a fixed price contract. All firms have the incentive to bid the true cost that they would incur given their type and first best effort levels.
Table 3.3: Rents, transfers and welfare of an optimal auction

<table>
<thead>
<tr>
<th>Number of Bidders</th>
<th>Rent</th>
<th>%</th>
<th>Transfer</th>
<th>Welfare Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,202,581</td>
<td>100.0</td>
<td>18,578,026</td>
<td>21,948,852</td>
</tr>
<tr>
<td>2</td>
<td>665,415</td>
<td>30.2</td>
<td>16,535,975</td>
<td>20,831,352</td>
</tr>
<tr>
<td>5</td>
<td>193,877</td>
<td>8.8</td>
<td>15,592,673</td>
<td>20,076,597</td>
</tr>
<tr>
<td>10</td>
<td>81,939</td>
<td>3.7</td>
<td>15,194,091</td>
<td>19,670,379</td>
</tr>
<tr>
<td>20</td>
<td>35,436</td>
<td>1.6</td>
<td>14,902,175</td>
<td>19,337,391</td>
</tr>
<tr>
<td>100</td>
<td>5,361</td>
<td>0.2</td>
<td>14,410,080</td>
<td>18,727,743</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>λ = 0.05</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,575,667</td>
<td>100.0</td>
<td>18,905,218</td>
<td>17,274,811</td>
</tr>
<tr>
<td>2</td>
<td>684,055</td>
<td>26.6</td>
<td>16,551,697</td>
<td>16,695,227</td>
</tr>
<tr>
<td>5</td>
<td>195,803</td>
<td>7.6</td>
<td>15,593,983</td>
<td>16,177,879</td>
</tr>
<tr>
<td>10</td>
<td>82,533</td>
<td>3.2</td>
<td>15,194,321</td>
<td>15,871,504</td>
</tr>
<tr>
<td>20</td>
<td>35,645</td>
<td>1.4</td>
<td>14,902,130</td>
<td>15,611,592</td>
</tr>
<tr>
<td>100</td>
<td>5,384</td>
<td>0.2</td>
<td>14,409,961</td>
<td>15,125,074</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>λ = 0.5</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,091,397</td>
<td>100.0</td>
<td>18,497,924</td>
<td>25,655,488</td>
</tr>
<tr>
<td>2</td>
<td>656,430</td>
<td>31.4</td>
<td>16,529,942</td>
<td>24,138,484</td>
</tr>
<tr>
<td>5</td>
<td>192,826</td>
<td>9.2</td>
<td>15,592,307</td>
<td>23,195,634</td>
</tr>
<tr>
<td>10</td>
<td>81,612</td>
<td>3.9</td>
<td>15,194,171</td>
<td>22,709,644</td>
</tr>
<tr>
<td>20</td>
<td>35,320</td>
<td>1.7</td>
<td>14,902,344</td>
<td>22,318,195</td>
</tr>
<tr>
<td>100</td>
<td>5,348</td>
<td>0.0</td>
<td>14,410,227</td>
<td>21,609,993</td>
</tr>
</tbody>
</table>

NOTES:

- Values are calculated for the average firm in the sample.
- The percentage column measures rents as a percentage of the total rents under an optimal contract (equivalent to n = 1).
Table 3.4: Rents, transfers and welfare of a Vickrey auction of fixed price contracts

<table>
<thead>
<tr>
<th>Number of Bidders</th>
<th>Rent</th>
<th>Transfer</th>
<th>Welfare Cost ($ \lambda = 0.3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>913,061</td>
<td>16,780,548</td>
<td>20,901,651</td>
</tr>
<tr>
<td>5</td>
<td>516,595</td>
<td>15,914,747</td>
<td>20,172,576</td>
</tr>
<tr>
<td>10</td>
<td>405,682</td>
<td>15,517,454</td>
<td>19,767,008</td>
</tr>
<tr>
<td>20</td>
<td>339,905</td>
<td>15,206,379</td>
<td>19,428,387</td>
</tr>
<tr>
<td>100</td>
<td>252,314</td>
<td>14,656,884</td>
<td>18,801,635</td>
</tr>
</tbody>
</table>

NOTES:

- Values are calculated for the average firm in the sample.
- Welfare cost was calculated using $\lambda = 0.3$.

As expected, the rents and transfer levels are slightly higher than under the optimal auction. However, it is interesting to note that the overall welfare cost levels are almost indistinguishable from the optimal auction. This result is not really that surprising. The auction serves to truncate from above the efficiency type distribution. Only relatively efficient firms will win the auction and for these firms the optimal contract is high powered, resembling a fixed price contract. However, confirming this result in an empirical context is still important. It implies that the tendering of fixed price contracts would be an excellent tool to achieve all the potential welfare benefits of the optimal auction. Furthermore, if bureaucratic or implementation costs are higher for the optimal auction, then the fixed price contract auction might be the best mechanism under an extended welfare criteria.

In order to arrive at the potential impacts of tendering on public finances we estimate, as above for the optimal contract, the transfer savings as a

\[ \text{This result still holds for other values of the cost of public funds.} \]
percentage of operational costs. For the average firm regulated through an
individually negotiated contract, transfers savings represent between 25.6% 
(when \( n = 2 \)) and 39.1% (when \( n = 100 \)) of operational costs. The equivalent 
figures for the standard cost contract are 23.9% and 38.2%.

### 3.6 Conclusions

We have estimated a structural model that enables different regulatory mech­
anisms to be compared. The mechanism that is potentially the most relevant 
from a policy perspective is the tendering of bus routes to the lowest bidder for a fixed transfer to service such routes. Our analysis shows that such 
a policy would probably achieve most of the welfare benefits of an optimal auction. Furthermore, we have estimated that the potential reduction in transfers from a fixed price auction could amount to between 24%-26% 
and 38%-39% of operational costs. In 1994 operational costs in the Norwe­
gian bus transport industry amounted to 2,689 million NOK. If our figures 
are correct, this implies that transfer savings for the regulatory authority of implementing an auction could be between 645 and 1,049 million NOK. In­cluding workshop labour and parts and materials, operational costs amounts to 3,504 million NOK. Applying the same figures to this measure we arrive at transfer savings of implementing an auction in the range of 841 to 1,367 million NOK.

The Norwegian state authority plans to reduce the block grant to the 
county authorities by approximately 700 million NOK when the full effect of tendering is materialised. If we assume that this figure is the authority’s estimated savings from the tendering process, it is on the lower end of the range of estimated savings according to our simulations.

It is interesting to compare our results with the experience of bus route 
tendering in London. According to Kennedy (1984), London Transport has estimated average cost savings, net of regulatory administrative costs, of 16%.
In other tendering experiences in the United Kingdom, cost savings have been estimated to be close to 20%\(^{29}\). Therefore, the range of cost savings we estimate for the Norwegian bus industry, although higher than these figures, are not unrealistic in comparison to the UK experience. In addition, there are several reasons to consider our results an optimistic upper bound on the potential savings from adopting a tendering approach to regulation. In the first place, there are probably political constraints that we cannot consider which limit the effects of an auction. For example, in order to avoid the collapse of inefficient firms, the authorities might introduce, implicitly or explicitly, some kind of quota system for parts of the network. Also, quality of service and long term fixed capital investment issues might preclude a pure tendering mechanism, such as the one we have considered in this paper. Finally, the benefits of tendering could also be reduced if the firms are able to collude in the bidding process.

Besides the numerical results and the policy simulations, the purpose of this paper has been to present a methodology for estimating cost functions that include adverse selection and moral hazard variables. This represents a first step to an empirical implementation of the new theory of regulation.

\(^{29}\)See Domberger, et al.(1986, 1987) for an analysis of the impact of competitive tendering of refuse collection services and hospital domestic services in the UK.
Table 3.5: Summary statistics of data for the Norwegian bus transport industry

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Mean of log</th>
<th>std</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>15651787</td>
<td>15.294</td>
<td>17579342</td>
<td>458137</td>
<td>111461911</td>
</tr>
<tr>
<td>Qnreg</td>
<td>298</td>
<td>5.110</td>
<td>349</td>
<td>10</td>
<td>1777</td>
</tr>
<tr>
<td>Qreg</td>
<td>1662</td>
<td>6.840</td>
<td>1617</td>
<td>75</td>
<td>6876</td>
</tr>
<tr>
<td>Capital</td>
<td>2107</td>
<td>7.205</td>
<td>1864</td>
<td>144</td>
<td>10826</td>
</tr>
<tr>
<td>Wage drvs.</td>
<td>112.455</td>
<td>4.714</td>
<td>14.810</td>
<td>62.417</td>
<td>178.348</td>
</tr>
<tr>
<td>Wage adms.</td>
<td>131.192</td>
<td>4.831</td>
<td>70.523</td>
<td>32.721</td>
<td>1474.964</td>
</tr>
<tr>
<td>Regime</td>
<td>0.450</td>
<td>—</td>
<td>0.498</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Share drvs.</td>
<td>0.751</td>
<td>—</td>
<td>0.052</td>
<td>0.549</td>
<td>0.867</td>
</tr>
<tr>
<td>Share adms.</td>
<td>0.137</td>
<td>—</td>
<td>0.047</td>
<td>0.029</td>
<td>0.319</td>
</tr>
<tr>
<td>Share fuel</td>
<td>0.112</td>
<td>—</td>
<td>0.028</td>
<td>0.015</td>
<td>0.210</td>
</tr>
<tr>
<td>Ratio Qu/Q</td>
<td>.184</td>
<td>0.145</td>
<td>0.016</td>
<td>.804</td>
<td></td>
</tr>
</tbody>
</table>

Firms = 88
years = 1987-1991

NOTES:

- The third column is the mean of the natural log of the variable for each observation (not the log of column two).
- Outputs are measured as vehicle kilometres (1000 unit-kms).
- Capital is the total number of seats of the rolling stock.
- Wages and costs are in 1991 NOK.
Table 3.6: Estimation results with alpha 10% higher

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>s.e.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.0644</td>
<td>0.5607</td>
<td>10.82</td>
</tr>
<tr>
<td>ln Capital</td>
<td>0.1869</td>
<td>0.0256</td>
<td>7.31</td>
</tr>
<tr>
<td>lnQ</td>
<td>0.8219</td>
<td>0.1531</td>
<td>5.37</td>
</tr>
<tr>
<td>lnQ^2</td>
<td>0.0042</td>
<td>0.0224</td>
<td>0.19</td>
</tr>
<tr>
<td>ln Wd</td>
<td>0.3951</td>
<td>0.0549</td>
<td>7.20</td>
</tr>
<tr>
<td>ln Wa</td>
<td>0.1032</td>
<td>0.0192</td>
<td>5.37</td>
</tr>
<tr>
<td>γ-1</td>
<td>-0.4459</td>
<td>0.0565</td>
<td>-7.90</td>
</tr>
<tr>
<td>b_in</td>
<td>0.5725</td>
<td>0.1063</td>
<td>5.39</td>
</tr>
<tr>
<td>b_sc</td>
<td>0.6314</td>
<td>0.1132</td>
<td>5.58</td>
</tr>
</tbody>
</table>

Firms = 88  
length of panel = 5  
Mean log-likelihood = 2.448

NOTES:

- Dependent variable is ln C.
- Homogeneity of the cost function has been imposed.
- Output is the simple sum of output in the regulated and unregulated market.
- alpha is 10% higher than in the base estimation.
Table 3.7: Estimation results with alpha 10% lower

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>s.e.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.0314</td>
<td>0.5527</td>
<td>10.91</td>
</tr>
<tr>
<td>In Capital</td>
<td>0.1865</td>
<td>0.0256</td>
<td>7.29</td>
</tr>
<tr>
<td>lnQ</td>
<td>0.8250</td>
<td>0.1505</td>
<td>5.48</td>
</tr>
<tr>
<td>lnQ^2</td>
<td>0.0038</td>
<td>0.0220</td>
<td>0.17</td>
</tr>
<tr>
<td>ln Wd</td>
<td>0.3946</td>
<td>0.0549</td>
<td>7.19</td>
</tr>
<tr>
<td>ln Wa</td>
<td>0.1031</td>
<td>0.0192</td>
<td>5.37</td>
</tr>
<tr>
<td>γ-1</td>
<td>-0.4463</td>
<td>0.0564</td>
<td>-7.91</td>
</tr>
<tr>
<td>b_in</td>
<td>0.5416</td>
<td>0.1088</td>
<td>4.98</td>
</tr>
<tr>
<td>b_sc</td>
<td>0.5974</td>
<td>0.1162</td>
<td>5.14</td>
</tr>
</tbody>
</table>

Firms = 88  
length of panel = 5  
Mean log-likelihood = 2.448

NOTES:

- Dependent variable is ln C.
- Homogeneity of the cost function has been imposed.
- Output is the simple sum of output in the regulated and unregulated market.
- alpha was 10% lower than in the base estimation.
Part III

Regulation: Auctions
Chapter 4

Tendering of refuse collection contracts in England

4.1 Introduction

This chapter concerns the tendering of refuse collection contract by English local authorities. In chapter 1, the advantages of auctions as a regulatory tool were discussed. It allows a regulator to reduce the rents earned by firms endowed with an informational advantage.

In 1988, compulsory competitive tendering (CCT) for refuse collection contracts was introduced in the United Kingdom. Refuse collection is a natural monopoly activity that needs to be regulated. Traditionally, the regulatory problem in this sector was solved by public sector provision. Prior to CCT, almost all local authorities delivered refuse collection services through an in-house team (the Direct Service Organisation (DSO)). Under this situation, regulation was probably akin to a cost plus contract where refuse collection costs were covered by local government budget outlays. Perhaps, a bargaining model between the DSO and the local authority, or between the local authority and central government, would be a more adequate description of the situation prior to CCT. The maintained hypothesis, however, is that before CCT there were scant incentives for efficiency improvements and the DSOs were earning some informational rents. If this was the case,
the introduction of competitive tendering should have reduced these rents and, consequently, the expenditure incurred by local authorities in refuse collection.

There is ample evidence that the tendering of contracts has reduced expenditures as theory would predict (Szymanski (1996)). However, there are doubts as to the fairness with which CCT has been applied. Close to 61% of CCT contracts (65% by value) have been won by DSOs (LGMB (1996)).

In this chapter we use an original data set on refuse collection contracts to explore certain issues related to CCT. In particular, a structural econometric model of auctions is used to estimate the underlying cost distributions of firms. We then test the proposition that the underlying cost distributions of DSOs and private firms are the same. If this proposition is accepted, it may suggest that there is no bias in the tendering process. Or, at least, that the high DSO success rate is not due to cost advantage of DSOs.

However, we find that DSOs that won a contract have lower costs than their private counterparts. The origin of this cost differential, and whether it is due to a bias in the tendering mechanism, is an open question. The results do have important policy implications which will be discussed later.

Besides providing an additional application of recent methodological advances in the econometrics of auctions, the topic of this chapter also has wider policy interest. The use of compulsory competitive tendering (CCT) for local authority blue collar services has been a controversial policy since it was introduced by the Conservative government in 1988 and the future of tendering as a regulatory instrument is in doubt. The Labour Party in the UK has pledged itself to abolishing CCT, claiming that competitive tendering should only be used as an instrument of last resort. As part of the preliminary data analysis, we show that tendering of contracts has a significant and negative effect on the expenditure incurred by local authorities on this service. Moreover, as theory predicts, the number of bidders for a given contract is negatively related to expenditure levels. These results cast doubts
on the wisdom of a policy aimed at abolishing CCT.

4.2 History of CCT

The Local Government Act introduced in April 1988, required that certain activities including refuse collection, street cleaning, building cleaning, schools and welfare catering, grounds maintenance and vehicle maintenance should be supplied under competitive conditions and laid down a timetable with dates by which specific services had to be subjected to competitive tendering.

Prior to the 1988 a reduced group of local authorities had voluntarily introduced competitive tendering for their refuse collection services. Besides the DSO, private firms were invited to bid for the provision of the service. There is strong evidence to suggest that these authorities spent less on this service compared to authorities serviced by a DSO (Domberger et al (1986), Cubbin et al (1987)).

Szymanski (1996), using data on 315 local authorities, found that CCT also had a significant effect on refuse collection expenditure. On average, expenditure per household was 19% lower after a local authority tendered their contract as compared to pre-CCT expenditure. In addition, 81% of local authorities in the sample registered at least some fall in cost.

Attributing the cost reductions to specific causes is more difficult. In general, costs reductions might come from one of three sources: lower wage payments to employees, lower standards of service or increased productivity. There is considerable anecdotal evidence to support the first explanation (LGIU (1994)). Prior to CCT many authorities maintained a policy of paying manual employees above the (often very low) market rate for the job. Paying higher than market wages is a concrete example of the firm, and it’s employees, earning rents under a low powered regulatory scheme. CCT forced all bidders to pay in line with market rates. Walsh (1991) also points
out that competition has affected fringe benefits, including loss of pension rights, sickness benefits and holiday pay.

There is no systematic evidence to support the view that cost reductions under CCT have on average been bought at the expense of service standards. Firstly, there is no evidence that authorities have lowered their specifications to reduce costs. Indeed, Szymanski shows that if anything, specifications have been raised in refuse collection. In general service standards were not well specified before CCT, and its introduction has obliged authorities to define their services in detail. Together with the separation of client-side and supplier functions implicit in CCT, this has led to a greater ability to monitor the performance of contractors, whether in-house DSOs or private firms. There is no evidence that contractors have been able to systematically deliver standards of service below those specified in the contract. A survey of 22,000 householders carried out by the Consumers' Association in 1994 (Consumers' Association (1995)) found that 86% were either "satisfied" or "very satisfied" with their service. Furthermore, there was no systematic difference in the level of satisfaction comparing local authorities where the DSO held the contract with authorities where it was held by a private contractor.

While there are no estimates of the effects of CCT on productivity, previous studies indicated that competitive tendering led to a significant improvement (see Cubbin et al (1987)).

Refuse collection is an attractive service to study mainly because of its simplicity and the homogeneity of the product across auctions. Inputs are basically unskilled labour and trucks, output is measured simply by the volume of waste collected and quality of service does not tend to vary widely because: either the garbage is taken away or it is not.
4.3 Data and descriptive statistics

The data on bidding numbers were collected by telephone questionnaire at the end of 1996. Local authority refuse collection client managers (who are responsible for the administration of the service) were asked how many bids had been received for their contract. The timetable to be followed by local authorities was imposed by central government and the first round of tendering under CCT rules took place between 1989 and 1993, so that in many cases contracts have now been re-tendered. We were unable to contact the relevant manager for some local authorities. In other cases client managers could not recollect or did not which to inform on the matter. However, responses were obtained for 175 contracts in England\(^1\). Of these, 113 where contracts tendered in the first round after CCT was introduced while the remaining 62 observations related to second round contracts.

Expenditure on refuse collection services were taken from data supplied to the Department of the Environment by each local authority. We focus here on the ex-post net expenditure by each local authority the first full year after the contract was tendered. This is the level of resources paid to the contractor. This variable, we assume, directly reflects the characteristics of the winning bid in the tendering process and will be, for the remainder of this chapter, considered as the “winning bid”.

Local authority characteristics included data on the number of domestic units (households) and non-domestic units (industrial and commercial) in each authority, size of the authority, rural or metropolitan. These last variables were also obtained from information provided by the Department of the Environment. A wage variable was created by linking the county level male full-time manual wage rate from the New Earnings Survey to each local authority. Characteristics also included political control, measured as the

\(^1\)One additional observation, the City of London, was not included in the data set given its highly peculiar structure and characteristics.
Table 4.1: Bidding and expenditure

<table>
<thead>
<tr>
<th>Number of bids</th>
<th>Number of authorities</th>
<th>Expenditure per household</th>
<th>Won by DSO (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>35.34</td>
<td>93.8</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>27.01</td>
<td>86.4</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>25.54</td>
<td>58.6</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>25.34</td>
<td>61.1</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>24.64</td>
<td>60.9</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>24.07</td>
<td>36.7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>24.55</td>
<td>50.0</td>
</tr>
<tr>
<td>8+</td>
<td>11</td>
<td>22.96</td>
<td>18.2</td>
</tr>
</tbody>
</table>

Notes:

- Expenditure was defined as the net expenditure on refuse collection expressed in April 1996 prices.
- * t-test for difference with category immediately above rejects equality of means at 5% confidence interval. The t-test is performed assuming unequal variances.

number of seats held by each political party. The data also included information on whether the contract was won by the in-house DSO and the contract start date taken from the CDC Contracts Handbook. The competitors to the DSO are usually private contractors, although there have been cases, albeit rare, where a DSO from a bordering authority has competed for a contract.

Other potentially relevant variables, such as whether the contract requires the winning firm to post a performance bond or the whether the authority made its depots, plants and machinery available to the contractor, could not be obtained. This is an unfortunate drawback of the data set².

²Hopefully future research will be able to overcome this limitation. The Local Government Management Board has a complete data set with detailed information on each contract tendered under CCT, including the number of bids received. However, at the moment it is not publicly available.
Table 4.1 gives the average expenditure per household on refuse collection services in the first full year following the competitive tender of a contract according to the number of bids received. The table shows that more bids are associated with lower expenditure. It also shows that the benefit of extra bids declines as number of bids increases. Expenditure per household was about 23.5% lower in authorities where there were two bids compared to authorities where there was only one. The expenditure per household was about 5.4% lower for authorities that received three bids compared to where there were two. The pattern continues for more bids, except for 7 bids. The Table also illustrates that the DSO was less likely to win when exposed to a higher degree of competition.

While Table 4.1 is indicative of the effect of having more bids, it is clear that in each local authority there are other factors which will affect the service expenditure. For example, large authorities may enjoy cost advantages over small ones. Also, there are potentially unobservable characteristics of local authorities which may affect both the number of bids submitted and expenditure.

CCT is a highly politicised issue and therefore differences in political control are likely to be important. Table 4.2 divides the sample into three groups, Labour controlled authorities, Conservative controlled authorities and others. Although, Labour controlled authorities tended to have a higher expenditure on the service before CCT, this difference was not significant at a 5% confidence level. After CCT, however, Conservative controlled authorities seem to have significantly lower expenditure levels. The differences in savings per household, however, are not statistically significant.

In order to control for these and other effects a regression approach is adopted below which allows us to isolate the relationship between bidding and costs holding all other factors constant.

\(^3\)However, the only statistically significant difference between contiguous categories is when the number of bids rises from 1 to 2.
Table 4.2: Differences related to political control

<table>
<thead>
<tr>
<th></th>
<th>Mean values</th>
<th>Differences in the means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Labour</td>
<td>Conservative</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>(n=57)</td>
</tr>
<tr>
<td><strong>Expenditure:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>per house pre-CCT*</td>
<td>32.73</td>
<td>30.40</td>
</tr>
<tr>
<td>per house post-CCT</td>
<td>28.41</td>
<td>23.52</td>
</tr>
<tr>
<td>Saving per household*</td>
<td>4.65</td>
<td>6.88</td>
</tr>
<tr>
<td><strong>Data on bidding:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bids received</td>
<td>3.18</td>
<td>4.49</td>
</tr>
<tr>
<td>Won by DSO (%)</td>
<td>0.46</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>Exogenous characteristics:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London (%)</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>Metropolitan (%)</td>
<td>0.26</td>
<td>0.12</td>
</tr>
<tr>
<td>Area (1,000's ha)</td>
<td>11.95</td>
<td>29.73</td>
</tr>
<tr>
<td>Domestic units (1,000's)</td>
<td>58.05</td>
<td>49.09</td>
</tr>
<tr>
<td>Other units (1,000's)</td>
<td>4.53</td>
<td>3.55</td>
</tr>
<tr>
<td><strong>Political Variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative share (%)</td>
<td>0.21</td>
<td>0.63</td>
</tr>
<tr>
<td>Labour share (%)</td>
<td>0.69</td>
<td>0.17</td>
</tr>
<tr>
<td>Total number of seats</td>
<td>51.21</td>
<td>48.23</td>
</tr>
</tbody>
</table>

**Notes:**

- All values in April 1996 prices.
- t-statistic is for differences in means assuming unequal variances.
- * Includes only 173 contracts for which past expenditure was available.
4.4 Preliminary data analysis: do the number of bids matter?

As a first step in the analysis of the data, we explore the relation between refuse collection expenditure and the number of bidders. This analysis has several purposes. First, it provides a general, albeit weak, test of auction theory. For most auction models, the winning bid (for a contract) will be a decreasing function of the number of bidders. Confirming this relationship in our data will show that it is at least consistent with theory. This adds some confidence that the data will be useful for the more ambitious structural estimation strategy. Second, there is a small empirical literature that looks at the behaviour of bids in a manner similar to ours. The analysis in this section adds another result to that literature.

4.4.1 Auction theory

What does the theory of auctions have to say regarding the relationship between winning bids and the number of bidders? First, we must define what bidding environment we are assuming. This environment encompasses not only the rules of the game (who wins and what he pays) but also the nature of the good being auctioned, the informational assumptions and the distribution of bidders’ valuation.

In CCT, we assume the lowest bid wins the auction and is paid the amount he bid. The process is basically a first price sealed bid auction with no reservation price, hence we will concentrate on this auction mechanism.

---

4See the review in chapter 1 and especially the discussion of section 1.6.2.
5McAfee and McMillan (1987) and Milgrom (1989) are good introductory surveys to auction theory.
6Notice that, to fit the characteristics of the phenomenon at hand, we couch the discussion in terms of agents bidding for the lowest price they are willing to receive to undertake a service. The underlying valuation of firms is the cost to the firm of providing the service.
7It is well known that a first price sealed bid auction is strategically equivalent to a Dutch auction. Therefore, the discussion below also applies to this last mechanism.
The assumption that local authorities choose the lowest bidder needs some further comment. CCT rules require the authority to award the contract to the most competitive bid. The Local Government Act of 1988 introduced the concept of anti-competitive behaviour to avoid an authority favouring its own DSO. Furthermore, under the 1988 Act the Secretary of State has the powers to take action against local authorities that breach these rules. These actions usually take the form of 'directions' requiring an authority to retender the contract or prohibiting the DSO from providing the service. Therefore, local authorities are under legal pressure to accept the most competitive bid.

With respect to bidders' costs (valuation), there are two polar cases, the common value and the private value paradigms. Under the private value paradigm, the cost to each firm, \( c_i \), which is known to the firm only, is independent of the costs of other firms. Competitors assume that the cost of firm \( i \) is a random draw from a common knowledge distribution \( F_i(c) \) with \( c \in [c, \bar{c}] \).

On the other hand, the common value paradigm assumes that each firm has the same unobserved cost, \( \bar{c} \), of providing the service. However, firms only observes a signal, \( s_i \), which are draws from the conditional distribution function \( F_i(s \mid \bar{c}) \) which has the property that for higher \( \bar{c} \) it is more likely to obtain a higher signal.

Which paradigm applies to a particular situation is hard to determine. Paarsch (1992) provides econometric tests to discern this issue. However, his approach is limited by the restrictive parametric assumptions which must be made in order to estimate the models. In most cases, researchers assume one or the other. In reality, all auctions will have a private and a common component and the affiliated model of Milgrom and Weber (1982) would be the most appropriate. However, some guidance as to the relative plausibility
of each paradigm can be provided by an examination of the case under study. For example, when the good being auctioned has a resale value (perhaps in the future) which is independent of the identity of the owner, and a big component of bidders valuation is this asset value, then the common value paradigm seems most appropriate. This is arguably the case for art objects and oil tracks.

On the other hand, if there is no resale value, and there is little uncertainty as to the valuation of the object, assuming private values may be reasonable. This, we believe, is the case with refuse collection contracts. There is no resale value to having a contract, the technology is simple and well known and accurate estimates of the costs of servicing an area should not be too difficult. In a survey of 220 private firms commissioned by the Department of the Environment, senior executives and managers did not mention, or show any concern, for uncertainty in cost estimates. If uncertainty as to the real costs of providing refuse collection services was important, we would have expected this to be an issue for firms. Therefore, cost differences between firms should more or less reflect their intrinsic efficiencies.

In the independent private values paradigm, if all bidders draw their valuation from the same distribution function (with identical support) then we have the symmetric private values model. The Bayesian Nash equilibrium bidding strategy (as a function of type $c_i$) for a first-price sealed bid auction is

$$b_i = c_i + \frac{\int_{c_i}^{\bar{c}} (1 - F(u))^{n-1} du}{(1 - F(c_i))^{n-1}}$$

where $b_i$ is the bid by firm that has a cost $c_i$, and $n$ is the number of bidders.

---

$^{10}$One main finding of the survey was that “Firms’ experience of operating contracts was generally very positive. Contractors had generally found little difficulty in meeting the requirements and conditions required for the contract in question” (BMRB (1995)).

$^{11}$Some contracts have clauses that index the payments made by local authorities. To the extent that these clauses transfer input price risks from the firm to the local authorities, they further reduce cost uncertainty for the firm.
Firms bid above their costs. Taking the derivative with respect to the number of bidders yields,

$$\frac{\partial b_i}{\partial n} = \int_{c_i}^{g} \ln \left[ \frac{(1 - F(u))}{(1 - F(c_i))} \right] (1 - F(u))^{n-1} du \leq 0$$  (4.2)

From (4.2) it is clear that the optimal bid for each firm is decreasing in $n$. Therefore, the expected winning bid must also be decreasing with the number of bidders.

As we have mentioned above, DSOs have won a disproportionate number of contracts. This casts doubts on the suitability of the symmetry assumption for modelling these tendering processes. We can relax the symmetry assumption by postulating a different (continuous) distribution, say $G(c)$, for the costs of one of the bidders (DSO). Then we have an asymmetric private values auction. Unlike the symmetric case, the equilibrium bidding functions can no longer be solved explicitly. We will defer the full presentation of the asymmetric model until section 4.6 below. For now, it suffices to state that bidding will also be more competitive in the asymmetric case as the number of bidders increases. Thus, expected expenditure on refuse collection should also decrease with more bidders.

In the case of the common value paradigm, optimal bids may eventually increase as the number of bidders increases. This is due to the influence of the winner's curse. Under a naive bidding strategy a bidder will win the auction when he had an overly optimistic signal $s_i$. Under symmetry, the winner's signal is lower than that received by all the other bidders. Therefore, the true cost of undertaking the contract is higher than he would have expected from his signal alone. Winning in this case is a curse because it indicates that the true cost is higher than expected. The winner's ex-post profits will be lower than expected or even negative. As the number of bidders increases, the winner's curse becomes more acute. Winning over more bidders is an even stronger indication that the signal was biased.
To avoid the winner’s curse bidders will optimally increase the bid to compensate for this effect\textsuperscript{12}. It could be that, as the number of bidders increases, this effect more than compensates the competitive effect on bidding and the resulting relationship between bids and number of bidders is U-shaped. However, as Laffont (1997) points out, there is no general proof for this result. It has been shown to hold only in some examples and simulated exercises.

However, even if individual bidding behaviour is U-shaped the expected winning bid will still be decreasing as more bidders are involved. This can be deduced from the result, proved in Wilson (1977) and Milgrom (1979), that as $n$ increases the expected selling price converges to the true value. In our data set we only observe the winning bid.

In summary, most typical auction models we predict a negative relationship between the number of bidders and the value of the winning bid. Therefore, our empirical analysis provides a weak test of the theory\textsuperscript{13}. If we do not find this negative correlation, then either the theory would be suspect, or the data has not been generated by a competitive auction mechanism.

### 4.4.2 Empirical specification

We specify the following function to examine the effects of bids on expenditure:

\[
\ln(E_{it}) = x_{it}'\beta + \alpha \ln(n_{it}) + \varepsilon_{it}
\]  

(4.3)

where $\ln(E_{it})$ is the logarithm of net expenditure on refuse collection in authority $i$ in the first full year after the contract has been tendered, $x_{it}$ is a

\textsuperscript{12}The key to determine the optimal bidding strategy is to note that the expected value of the object (or cost) conditional on winning is not the same as the unconditional expected value. The optimal strategy should maximise expected profits conditional on winning.

\textsuperscript{13}Other authors have analysed very similar issues in alternative contexts. See, for example, Gilley and Karels (1981), Brannman, Klein and Weiss (1987) and Giliberto and Varaiya (1989).
A second objective of this section is to explore the factors that determine whether the DSO wins the contract. As discussed above, we assume that once tenders are submitted, the contract is awarded to the lowest bid\(^\text{15}\). Therefore, the probability of the DSO winning the contract will also be a function of the number of bids submitted. We will model this using a latent variable approach. We define the continuous variable \(y^*\) as the DSO’s cost advantage over competitors. This variable will depend on some exogenous factors \(z\), the number of bids \(n\) and some unobservable stochastic effect \(\varepsilon\),

\[
y_{it}^* = z_t' \gamma + \delta \ln(n_{it}) + \mu_{it}
\]

where, as before, \(\gamma\) and \(\delta\) are conformable parameter vectors.

If the DSO submitted a lower bid than the others, \(y^*\) is positive. We do not observe the variable \(y^*\), only the indicator variable \(y\) which takes a value of one if the DSO wins the contract and zero otherwise. Our observation rule for \(y\) is,

\[
y = \begin{cases} 
1 & \text{if } y^* \geq 0 \\
0 & \text{if } y^* < 0
\end{cases}
\]

This specification leads naturally to a probit estimation of \(\gamma\) and \(\delta\) if one is willing to assume a normal distribution for \(\mu\).

How do we interpret \(\varepsilon\) and \(\mu\) in equation (4.3) and (4.4)? In the case of \(\varepsilon\), it could represent random shocks that influence expenditure, but it will also

\(^{14}\)The relation between the winning bid and \(n\) will probably be highly non-linear. In the empirical implementation we experimented with powers of \(\ln(n)\) to capture possible non-linearities.

\(^{15}\)As will be seen later, the structural analysis casts some doubts on this assumption. It would seem that local authorities are choosing the winner based on characteristics other than expenditure.

\(^{16}\)Once again, theory would predict that the number of bidders is related in a non-linear way to \(y^*\).
include unobservable (to the econometrician) variables that are correlated with bids. One example could be the local authority's attitude to competitive tendering and private provision of refuse collection services\textsuperscript{17}. Furthermore, informal discussions with participants reveals that private refuse collection companies are aware of these attitudes, and their decision to participate in a particular tendering process will depend on the perceived willingness of a local authority to allow private provision of the service. This implies that the number of bidders for a particular contract may be endogenous and correlated with the residual \( \varepsilon \), leading to biases in the estimated parameters.

Besides attitudes to CCT, there may be other characteristics of a local authority that influences both expenditures and the number of bidders but which is not accounted for by the observable variables contained in our data set. For example, geographical characteristics, distances and availability of refuse disposal sites, or any other authority specific effect.

For the above reasons, we assume that the disturbance term \( \varepsilon \) has an error component structure given by

\[
\varepsilon_{it} = \tau_i + v_{it} \tag{4.5}
\]

where \( \tau_i \) is an authority specific effect, potentially correlated with the number of bids and other right hand side variables, and \( v_{it} \) is random disturbance uncorrelated across time and across authorities.

Similar arguments can be presented to indicate that the number of bidders may also be correlated with \( \mu \). If local authorities can design the contract in such a way that the local DSO can present a lower bid, then the probability of the DSO winning will be higher. But at the same time, the number of companies willing to spend the necessary time and resources to present a bid

\textsuperscript{17}Some commentators believe that local authorities can influence the outcome of the bidding process by designing the contract in such a way that it favours the DSO. On the other hand, others have mentioned that private companies show an unfounded lack of interest in bidding for contracts in authorities controlled by a particular political party.
will be lower. We define the structure of $\mu$ therefore to be

$$\mu_t = \zeta_t + \varphi_t$$  \hspace{1cm} (4.6)

with the interpretation of each term analogous to (4.5) above.

The assumed error structures implies that standard OLS and Probit estimates of equations (4.3) and (4.4) will be biased if there is some correlation between the authority specific effect, $\tau_i$, and $\zeta_i$, and any of the regressors, including the number of bids. In the estimation strategy below, we take account of this potential endogeneity bias.

4.4.3 Results

Table 4.3 and 4.4 presents the results for the expenditure equation and the probit equation, respectively. The dependent variable in the regressions of table 4.3 was the logarithm of net expenditure on refuse collection the first full year after the tendering of the contract (in April 1996 prices). The conditioning variables were the logarithm of the real wage rate (regional wage from the New Earnings Survey), the logarithm of the number of domestic units in the local authority's jurisdiction, the logarithm of the number of non-domestic units, and the logarithm of bids received. Other potential regressors such as size of the local authority, a trend, and the metropolitan and London dummy variables were found to be insignificant and were not included in the regressions shown below. Another important influence on costs should be fuel and motoring costs. The use of the RPI motoring expenses subindex was found to be insignificant and is also not included in the results below. A potential explanation for this last result is that there is not enough variation across time in the RPI index to identify the effects of motoring costs. Since, the RPI subindex is a national average there is no cross section variation in this variable and identification must rely on the time series variations of this variable. However, the time period under consideration is only 5 years long.
Table 4.3: Expenditure on refuse collection regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>-3.1602</td>
<td>-2.9689</td>
<td>-4.3202</td>
<td>-2.1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.45)</td>
<td>(-2.31)</td>
<td>(-3.23)</td>
<td>(-1.75)</td>
</tr>
<tr>
<td>Ln(wage)</td>
<td></td>
<td>0.1376</td>
<td>0.1092</td>
<td>0.2843</td>
<td>0.1375</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.57)</td>
<td>(0.45)</td>
<td>(1.18)</td>
<td>(0.63)</td>
</tr>
<tr>
<td>Ln(domestic)</td>
<td></td>
<td>0.9379</td>
<td>0.9310</td>
<td>1.0271</td>
<td>0.5981</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.21)</td>
<td>(12.12)</td>
<td>(11.98)</td>
<td>(5.90)</td>
</tr>
<tr>
<td>Ln(non-domestic)</td>
<td></td>
<td>-0.0469</td>
<td>-0.0334</td>
<td>-0.1088</td>
<td>-0.0726</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.65)</td>
<td>(-0.47)</td>
<td>(-1.43)</td>
<td>(-1.06)</td>
</tr>
<tr>
<td>ln(n)</td>
<td></td>
<td>-0.1988</td>
<td>-0.5163</td>
<td>-0.1794</td>
<td>-0.1368</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-6.69)</td>
<td>(-2.79)</td>
<td>(-5.69)</td>
<td>(-4.71)</td>
</tr>
<tr>
<td>ln(n)^2</td>
<td></td>
<td>-</td>
<td>0.2746</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(n)^3</td>
<td></td>
<td>-</td>
<td>-0.0623</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-1.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol. Control_{con}</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-0.3015</td>
<td>-0.2175</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-3.10)</td>
<td>(-2.46)</td>
</tr>
<tr>
<td>Pol. Control_{lab}</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-0.0989</td>
<td>-0.0721</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-1.18)</td>
<td>(-0.96)</td>
</tr>
<tr>
<td>Ln(Cost)_{t-1}</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.4039</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.34)</td>
</tr>
</tbody>
</table>

\[ Adj \ R^2 \quad 0.84 \quad 0.84 \quad 0.85 \quad 0.87 \]

\[ Number \ of \ obs. \quad 174 \quad 174 \quad 174 \quad 172 \]

Notes:

- OLS estimates. t-ratios in parentheses. Dependent variable was the logarithm of net expenditure on refuse collection the first full year after the contract started in April 1996 prices.

- Test statistic for the null hypothesis that parameters on \( \ln(n)^2 \) and \( \ln(n)^3 \) are jointly equal to zero in the second regression was \( F(2, 167) = 1.86 \).
The third and fourth column include political variables as regressors. The variables are the fraction of seats controlled by the Conservative Party and the Labour Party in the local government, respectively. The justification for including these variables is to control for unobserved attitudes towards competition that might affect costs. In other words, in this specification the political control variables are assumed to be proxies for the authority specific effect $\tau_i$. Still another specification uses the expenditure level of the year before the local authority tendered the contract.

The wage rate coefficient is positive, as expected, but it is not estimated very precisely. The most important determinant of costs is the number of domestic units served\textsuperscript{18}. In the first three regressions, the null hypothesis of no economies of scale in domestic units served cannot be rejected. In the last regression, there is no evidence against long-run constant returns to scale.

The number of bids has a significant and negative impact on costs. This relationship seems to be linear as the higher order terms are not significant. An F-test does not reject the null hypothesis that the coefficients on the higher order bid variables are jointly equal to zero. Including the political variables does diminish the coefficient on costs but it still remains significant and negative.

Using the coefficient on bids from the first and third regression gives an estimated expenditure savings shown in table 4.5. The values of table 4.5 were calculated as

$$C = kn^a$$

where $k$ has been normalised to equal 100. As bids increase, expenditure falls by less than the unconditioned data of table 4.1 would suggest, although strict

\textsuperscript{18}Previous studies of refuse collection costs have found similar results. Domberger, et al (1986), using a richer data set than ours, find that other output characteristics such as frequency of service and method of collection do affect costs. However, the added explanatory power provided by these other variables is very low. Most of the variation in costs across authorities seems to be explained by units served.
Table 4.4: Probability of DSO winning contract

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>1.5922</td>
</tr>
<tr>
<td></td>
<td>(5.19)</td>
</tr>
<tr>
<td>ln(n)</td>
<td>-0.9975</td>
</tr>
<tr>
<td></td>
<td>(-4.81)</td>
</tr>
<tr>
<td>ln(n)^2</td>
<td>-0.7419</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
</tr>
<tr>
<td>ln(n)^3</td>
<td>-0.2332</td>
</tr>
<tr>
<td></td>
<td>(-0.56)</td>
</tr>
<tr>
<td>Pol. Control_{con}</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol. Control_{lab}</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Date trend</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.12</td>
</tr>
<tr>
<td>Num. of obs.</td>
<td>175</td>
</tr>
</tbody>
</table>
comparisons cannot be made since the data of table 4.1 is expenditure per household. The presence of two bidders reduces expenditure by about 12% to 13% compared with the single bidder case. The average savings from CCT over the sample is between 20% to 22%, consistent with previous studies on the subject (Domberger, et al (1986); Szymanski (1996))^19.

The Probit estimations show that increasing bids lowers the probability of the DSO winning the contract. The date trend variable, measured as the number of days elapsed from 1 January 1988 to the starting date of the contract, is negative and significant. This implies that as CCT evolved, the DSOs were less likely to win a contract.

In all, it would seem that the CCT tendering of refuse collection contracts has generated competitive pressures that serve to lower expenditure. This accords well with the theoretical predictions from auction theory. However, as was discussed above, the number of bids may not be exogenous. Contractors may decide to bid for a particular contract depending on the characteristics of

---

Notice that because costs and the number of bids have a non-linear relationship in our specification, the average number of bids, which is 4.3 in the sample, is not the correct variable to use when calculating expected savings. Expected savings are given by $E[N\alpha]$ not $E[N] \alpha$. 

---

$\begin{array}{|c|c|}
\hline
\text{Number of bids} & \text{Cost} \\
\hline
1 & 100.0 \\
2 & 88.3 \\
3 & 82.1 \\
4 & 78.0 \\
5 & 74.9 \\
10 & 66.2 \\
\hline
\end{array}$

---

19 Notice that because costs and the number of bids have a non-linear relationship in our specification, the average number of bids, which is 4.3 in the sample, is not the correct variable to use when calculating expected savings. Expected savings are given by $E[N\alpha]$ not $E[N] \alpha$. 

---

200
the local authority, especially the willingness of a local authority to contract the service to a private firm\textsuperscript{20}. This attitude of a local authority will probably be correlated with the vigour with which the local authority pursues cost efficiency in refuse collection, which in turn may be part of the residual of the expenditure function. Including political variables in the expenditure equation, as was done earlier, controls for part of these effects. However, there might be some additional factors which are not captured by the political control variables.

A more robust estimating strategy is to first difference equation (4.3)

$$ln(E_{it})-ln(E_{i(t-1)}) = (x_{it}-x_{i(t-1)})'\beta + \alpha (ln(n_{it})-ln(n_{i(t-1)}))+\varepsilon_{it}-\varepsilon_{i(t-1)}(4.7)$$

which will eliminate the authority specific effect \(\tau_i\) (as well as any other time invariant variable). In the above first differenced model, lagged variables are defined as the values in the year before the particular contract was tendered. For those observations that correspond to the first wave of contracts tendered under CCT, the lagged variables are for a year when the service was provided by the DSO under monopoly conditions. Therefore, \(n_{i(t-1)}\) could be assumed to be 1 for these authorities, making the logarithm zero. For second round observations, \(n_{i(t-1)}\) should be the number of bids that each authority received for the first round contract. Equation (4.7) can then be written as

$$\Delta lnE_i = \Delta \bar{x}'\beta + \alpha ln(n_{it}) - \alpha 1(round = 2)ln(n_{i(t-1)}) + \nu_{it} - \nu_{i(t-1)} (4.8)$$

where \(\bar{x}\) contains only the time varying variables of the original vector \(x\), and \(1(round = 2)\) is an indicator function which takes a value of one if the observation is a second round contract. We do not have information on the past number of bids for second round contracts, therefore, the whole term \(1(round = 2)ln(n_{i(t-1)})\) is proxied by a dummy variable that takes a value of

\textsuperscript{20}Evidence from the BRMB (1995) survey states that "Most firms had been selective about the contracts for which they bid, choosing contracts on the basis of geographical location, size of contracts and mix of work".

201
one if the observation is a second round contract. Notice that because $\alpha$ is expected to be negative, the coefficient on the second round dummy should be positive. Expenditure reductions for second round contracts should be smaller than for first round contracts since part of the monopoly rents would already have been eroded when the first contract was tendered.

Table 4.6 shows the results from the regression on the first difference model. For completeness the first column shows the estimation omitting the second round dummy. The results show that the number of bids did affect the change in expenditure achieved through tendering, however the parameter in our preferred specification (column 2) is reduced by 50% from the undifferenced model.

Column 3 indicates that there is no evidence of a non-linear relationship between expenditure reductions and the number of bids. The F-statistic for the hypothesis that the parameters on $\ln(n)^2$ and $\ln(n)^3$ are jointly zero is $F(2, 163) = 1.30$, which is not rejected at conventional significance levels. Also, as expected, second round contracts achieved less expenditure savings than first round contracts. The fourth regression shows that the units variables ($\ln(dominic)$) as well as the political variables are insignificant in explaining the change in refuse collection expenditures. This is to be expected if the authority specific fixed effects model is correct.

Finally, it could be argued that there is still some additional time varying unobservable effect that might be correlated with the number of bids. First differencing the model might not eliminate all the potential endogeneity as bids could still be correlated with $(v_t - v_{t-1})$. This could be the case, for example, if the attitude to CCT of a local authority changed during the period before the contract was awarded and the time the contract was tendered. The fifth column of table 4.6 shows the results of an instrumental variable regression on the first differenced model. All variables were assumed to be exogenous except for the number of bids. This last variable was instrumented by a date trend, the political control variables, and the metropolitan
Table 4.6: Regressions of first differenced model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0794</td>
</tr>
<tr>
<td></td>
<td>(-1.78)</td>
</tr>
<tr>
<td>Δln(wage)</td>
<td>0.5700</td>
</tr>
<tr>
<td></td>
<td>(0.99)</td>
</tr>
<tr>
<td>ln(n)</td>
<td>-0.0813</td>
</tr>
<tr>
<td></td>
<td>(-2.58)</td>
</tr>
<tr>
<td>ln(n)^2</td>
<td>0.3105</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
</tr>
<tr>
<td>ln(n)^3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.62)</td>
</tr>
<tr>
<td>ln(domestic)</td>
<td>-0.0281</td>
</tr>
<tr>
<td></td>
<td>(-0.69)</td>
</tr>
<tr>
<td>Pol. Control&lt;sub&gt;Con&lt;/sub&gt;</td>
<td>-0.1106</td>
</tr>
<tr>
<td></td>
<td>(-1.15)</td>
</tr>
<tr>
<td>Pol. Control&lt;sub&gt;Lab&lt;/sub&gt;</td>
<td>-0.0530</td>
</tr>
<tr>
<td></td>
<td>(-0.64)</td>
</tr>
<tr>
<td>Scnd. round</td>
<td>-0.1077</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
</tr>
</tbody>
</table>

| attr. R^2               | 0.03         | 0.07         | 0.07         | 0.07         | -            |
| Number of obs.           | 169          | 169          | 169          | 169          | 169          |

Notes:
- The dependent variable was the first difference of the logarithm of net real expenditure on refuse collection. All values in April 1996 prices.
- The information on past expenditure and past wages were available for only 169 of the 175 observations.
- regression 5 is a 2SLS estimation with date trend, political variables, and the geographic dummies used as instruments for the number of bids. Standard errors have been calculated using corrected variance estimator.
and London dummies.

The justification for using the trend variable as an instrument is multiple. First, there might be a learning process for contractors in bidding or interest in CCT might have grown as the program developed. In addition, entry and mergers in the industry will have had an effect in the number of companies operating and, therefore, the potential number of bidders. It is important to note that the Department of the Environment divided local authorities into 6 groups and established a staggered statutory timetable for the tendering of contracts for each group. Authorities were selected into each group according to alphabetical order and therefore it is reasonable to argue that the dates at which contracts were tendered were not correlated with local authority characteristics. This would render the trend variable a valid instrument. The London and metropolitan variables may also be correlated with the number of bidders if the number of potential firms varies by geographical area. In the case of these last variables, however, the case for them being valid instruments is much weaker.

Results of column 5 show that the original conclusion that bids reduce expenditure is not reversed. In fact, the coefficient on bids is actually larger in absolute value compared to column 2.

Next we perform several specification tests on the above estimations. Under the null hypothesis that the individual local authority effects are not correlated with the regressors, then the parameters estimated using the expenditure equation in levels, equation (4.3), would be consistent. The estimates from the first differenced model would also be consistent, albeit less efficient than the previous estimates. Under the alternative hypothesis of correlation between the individual effects and some regressors, the estimates using equation (4.3) are inconsistent while the first differenced results are still

21Using a year trend, although less precise, gives the same result as the trend variable defined in terms of days.
22In the regressions of table 4.3 these variables were not significant. However, they are significant in a regression on the number of bids.
consistent. Therefore, we can test the hypothesis of correlated fixed effects
by comparing both sets of estimates.

To be more precise, assume the following model

\[ y_{it} = \beta_0 x_{it} + \varepsilon_{it} \]  \hspace{1cm} (4.9)

where the vector \( x_i \) contains regressors potentially correlated with \( \varepsilon_{it} \), and \( \beta_0 \) is the true parameter vector. Under the null of no correlation between \( x_{it} \) and \( \varepsilon_{it} \), an OLS estimate of equation (4.9) will yield consistent estimates, say \( \hat{\beta}_1 \). Estimates using a first differenced model (or in another context an instrumental variable approach) will yield estimates \( \hat{\beta}_2 \). Under the null, \( \text{plim}(\hat{\beta}_2 - \hat{\beta}_1) = 0 \), since as the sample size grows the two methods of estimation should yield the same estimates of the true parameter vector \( \beta_0 \). Under the alternative, however, the two estimates may no be equal.

One way to test the null hypothesis of no correlation is to perform an \( \chi^2 \) test of the difference \( (\hat{\beta}_2 - \hat{\beta}_1) \). If \( \hat{\beta}_1 \) is efficient, then the covariance matrix of the vector of differences is very easy to compute (Hausman (1978)). However, in our context, \( \hat{\beta}_1 \) will not necessarily be an efficient estimator and the covariance matrix is more involved (Mroz (1987)).

First we tested the null hypothesis that the regressors are not correlated with the individual effects. Call \( \beta_{OLS} \) the vector of parameter estimates for the wage rate and the number of bids variable from the levels equation and \( \beta_\Delta \) the estimates from the first difference model\(^{23}\). Our preferred specification in the first difference model is regression 2 of table 4.6 while for the levels model it is regression 1 of table 4.3. The hypothesis of equal parameters is clearly rejected by the results shown in the first row of table 4.7. This implies that authority specific fixed effects seem to be present and, moreover, they are correlated with some right hand side variables in the levels equation.

\(^{23}\)Since any time invariant regressors are not identified from a first differenced model, the test are performed only for the wage rate and the number of bids variable.
### Table 4.7: Hausman specification tests

<table>
<thead>
<tr>
<th>Test results</th>
<th>X ~ χ²(2)</th>
<th>P(x&gt;X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) $\beta_\Delta - \beta_{OLS}$</td>
<td>9.29</td>
<td>0.0096</td>
</tr>
<tr>
<td>b) $\beta_{IV} - \beta_\Delta$</td>
<td>0.63</td>
<td>0.7303</td>
</tr>
</tbody>
</table>

- Both tests are for the equality of coefficients on the bids and wage variables.

- $\beta_\Delta$ is the vector of coefficients from regression 2 of table 4.6. $\beta_{OLS}$ is the vector of coefficients from regression 1 of table 4.3. $\beta_{IV}$ is the vector of coefficients from the instrumental variable regression 5 of table 4.6. It an instrumental variable estimation on the first differenced model.

- The covariance matrices of the estimators were computed as shown in the appendix of Mroz (1987). They are robust to arbitrary heteroscedasticity.
The other test compares the estimates of the column 5 of table 4.6 with column 2. The null hypothesis would be that there is no correlation between the differenced regressors and the differenced error term. The second row of table 4.7 shows that there is not much evidence against parameter equality. Time varying disturbances do not seem to be correlated with the regressors.

We also tested for exogeneity in the probit equation (4.4) using the Blundell-Smith test (Smith and Blundell (1986)). This test is a generalisation of the Wu-Hausman test to a limited dependant variable context. It consists of running a reduced form regression on the number of bids using all exogenous variables and instruments as regressors. The fitted residuals of this regression are then included in the probit equation as an additional regressor. The Blundell-Smith test is a t-test on the coefficient of this residual variable in the probit equation. Under the null hypothesis of exogeneity, this variable should be insignificant, which is the case for the results in table 4.8. There is not much evidence that the number of bids is endogenous. Furthermore, the coefficient on bids is more negative than before, although it is not significantly different from zero in the second specification.

4.4.4 Conclusions of preliminary analysis

The introduction of Compulsory Competitive Tendering in the UK has generated a natural experiment for testing the relationship between bidding and expenditures on refuse collection services. We find that a greater number of bids submitted is associated with a lower expenditure levels. However, if we control for authority specific fixed effects, the impact of bidding, although still significant, is below that implied by the unconditional data or by regression analysis that does not take into account the potential correlation between the number of bids and the fixed effect.

Table 4.9 shows the implied expenditure savings using the bids coefficient of the first differenced model. For the case of 4 bids, the cost saving would be
Table 4.8: Probit estimation of DSO success rate with endogeneity correction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>2.9567</td>
</tr>
<tr>
<td></td>
<td>(5.15)</td>
</tr>
<tr>
<td>$\ln(N)$</td>
<td>-1.5346</td>
</tr>
<tr>
<td></td>
<td>(-3.67)</td>
</tr>
<tr>
<td>Pol. Control$_{con}$</td>
<td>-1.6029</td>
</tr>
<tr>
<td></td>
<td>(-2.83)</td>
</tr>
<tr>
<td>Pol. Control$_{lab}$</td>
<td>-1.1230</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Trend</td>
<td>-0.0403</td>
</tr>
<tr>
<td></td>
<td>(-2.79)</td>
</tr>
</tbody>
</table>

| Blundell/Smith Statistic | 1.87 | -0.84 |
| Number of obs.           | 175  | 175   |

Notes:

- Probit estimates. t-ratios in parentheses. Standard errors were not corrected for generated regressor.

- Instruments for bids variable in column 1 were the political control variables, and the London and metropolitan dummies. For column 2 the instruments were the London and metropolitan dummies.

- Blundell/Smith Statistic is distributed asymptotically as a standard normal.
Table 4.9: Cost reduction effects of the number of bids implied by the first differenced model.

<table>
<thead>
<tr>
<th>Number of bids</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>93.3</td>
</tr>
<tr>
<td>3</td>
<td>89.6</td>
</tr>
<tr>
<td>4</td>
<td>87.0</td>
</tr>
<tr>
<td>5</td>
<td>85.1</td>
</tr>
<tr>
<td>10</td>
<td>79.4</td>
</tr>
</tbody>
</table>

about 13% of the original level. Although significant, it is below the 22%-24% range implied by the results of table 4.3. An explanation for the discrepancy is that some local authorities have lower costs due, perhaps, to some special characteristic of the authority or an attitude towards the efficiency with which local government services should be provided. In turn, these special characteristics may have induced a higher number of firms to bid for the refuse collection contract. The bottom line, however, is that even correcting for the potential endogeneity of the number of bids, competition seems to have a significant impact in reducing expenditures by local authorities.

4.5 Structural analysis

In the case of refuse collection, one stylised fact of CCT is that the DSOs have won a disproportionate number of contracts. As already mentioned, information from the Local Government Management Board shows that 61.3% of tendered contracts have been awarded to DSOs (LGMB (1996)). In our data set, the comparable figure is 59.4%.
Table 4.10: Binomial test for DSO success rate

<table>
<thead>
<tr>
<th>Number of bids (n)</th>
<th>Number of contracts (t_n)</th>
<th>Won by DSO ((\hat{p}_n))</th>
<th>Expected proportion ((\frac{1}{n}))</th>
<th>Probability ((p_n &gt; \hat{p}_n))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>22</td>
<td>19</td>
<td>0.50</td>
<td>0.0004</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>17</td>
<td>0.33</td>
<td>0.0045</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>22</td>
<td>0.25</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>14</td>
<td>0.20</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>11</td>
<td>0.17</td>
<td>0.0067</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>4</td>
<td>0.14</td>
<td>0.0180</td>
</tr>
</tbody>
</table>

Notes:

- Last column is the probability (from a binomial distribution) of obtaining a DSO success rate higher or equal to the one observed.
- Calculations were not performed for auctions with more than 7 bidders since the number of such auctions in each category was small.

If the tendering mechanism was fair and local authorities were only concerned about refuse expenditure, the firm with the lowest bid would win. In a private value auction with symmetric bidders, equilibrium bidding strategies are identical for all firms and given by equation (4.1). This bidding function is strictly increasing on costs. Therefore, the lowest bidder will be the firm with the lowest cost. The probability of a given firm (firm i) winning the auction would be given by

$$P[i \text{ wins}] \int_{x}^{e} f(v) [1 - F(v)]^{n-1} dv = \frac{1}{n}.$$ 

Therefore, the number of times that the DSO wins when there are \(n\) bidders, \(\hat{p}_n\), will be distributed as a binomial with probability parameter, \(\frac{1}{n}\), and parameter \(t_n\). This last parameter is the number of auctions where \(n\) bidders are observed.
Table 4.10 shows the results of testing the proportion of DSO successes against the expected value under the null hypothesis of symmetry and fairness in the bidding mechanism. There is overwhelming evidence to reject the null hypothesis of symmetry and fairness. Except for auctions where 7 bidders participated, the hypothesis can be rejected at a 1% confidence level.

The natural question that arises is what is the exact cause of the high DSO success rate. There are several hypothesis that may explain this phenomenon.

First, there may be a lack of interest of private firms to bid for contracts in certain areas. A survey of private contractors carried out by BMRB on behalf of the Department of the Environment shows that “Most firms had been selective about the contracts for which they bid, choosing contracts on the basis of geographical location, size of contract and mix of work” (BMRB (1995)). However, this explanation does not fit well with the proportions test carried out above. Although, there are many contracts for which there were low numbers of bidders, the DSO success rate is still too high to be explained by this factor. Alternatively, it may be that private firms do not always present serious bids. But, given that there are some costs of preparing and presenting a tender, there is really no compelling reason why firms would behave in this manner.

Second, DSOs may have a cost advantage over private firms. If true, it would imply that the auctions are not symmetric. In this section we want to explore the plausibility of this hypothesis.

One source of cost advantage may be geographical. DSOs are based inside the local authority jurisdiction, while private firms may have to incur extra motoring costs due to the need to transfer trucks and employees from other localities. Other incumbency advantages may also lower DSO costs relative to private firms.  

24 Many authorities allow private contractors to use the facilities and depots of the DSO if they win the contract. This would reduce any geographical cost advantage of the DSO.  
25 Notice that by assuming that CCT contracts can be modelled as private value objects, we are ruling out any informational advantage that the DSO may have with respect to the
Another source of cost advantage may be that some local authorities are biasing the tendering mechanism by making private firm's costs artificially higher as compared to the DSOs. Possible candidates for the source of bias are the performance bonds that private firms (but not DSO's) have to post in some local authorities, the complexity of the tender documents, or the design and specifications of the contract itself.

Conversations by the author with local authority officials indicates that the contract design may be an important factor. Stipulating capital requirements (number and type of trucks, etc.) which are close to those held by the DSO would give a cost disadvantage to private firms that do not have the exact capital type and mix and, consequently, would require further investments to meet contract specifications.

The Department of the Environment (DoE) seems to have implicitly recognised a cost advantage to the DSO. As a way to avoid unfair competition from the use of publicly owned assets, the DSO is required to break even in a financial year after allowing for a return on these assets. The DoE raised the rate of return over assets that DSOs must earn from 5% at the beginning of CCT, to 6% in 1994. This should raise the revenue required by DSOs and consequently the value of bids submitted by these organisations.

We propose to examine the plausibility of the relative cost hypothesis by estimating the underlying costs of each type of firm. Specifically, we use a non-parametric estimation procedure for private value auctions due to Elyakime, Laffont, Loisel and Vuong (1994) and Guerre, Perrigne and Vuong (1995a), and extended by Laffont and Vuong (1997), to recover the underlying cost of winning firms. The maintained assumption is that observed bids are equilibrium outcomes of asymmetric auctions and that the lowest bidder wins the auction.

cost of servicing the contract. If this were an important consideration, then a common value auction model would be more appropriate than the private value model we are assuming. However, see the discussion in section 4.4.1.
The estimated cost distributions are informative about the potential cause of the high DSO success rate. Although, for reasons discussed below, the results must be treated with great care, we find that, on average, DSOs that have been successful in winning a contract have lower costs than successful private firms. This may indicate either of two things. That DSOs have an incumbency advantage or that there are elements in the tendering process that give a DSO a cost advantage over private firms.

Besides exploring the sources of asymmetry in CCT auctions, the exercise undertaken below is a first step towards the estimation of the whole distribution of costs for each firm. Recovering the underlying cost distributions has important policy implications. Myerson (1981) showed that if bidders are asymmetric, the optimal revenue maximising mechanism is not one of the common auctions observed in practice: English ascending price, Dutch descending price, first price sealed bid, or Vickrey auction. In fact, if the distributions are the same except for their means, a revenue maximising (or expenditure minimising) auctioneer should bias the auction in favour of the less efficient type (McAfee and McMillan (1987)).

Unlike other structural estimation methods proposed for auction data, the approach used in this paper has the advantage that we do not have to assume a parametric family for the distribution of costs. This is very important in the context of asymmetric auctions. In this case, the relative performance of auction mechanisms depends on the shape of the distributions. It is possible that an English auction raises more or less revenue (in expectation) than a first price auction depending on the form of the distributions (Maskin and Riley (1996)). Making parametric assumption about the underlying costs. What is estimated below is the distribution of winning costs for each type of firm.

Expenditure minimisation may not be a suitable objective for a local authority, however. The optimal auction awards the contract to an inefficient firm with positive probability and therefore is not always Pareto optimal.
form of the distribution would obviously make comparisons between auction forms spurious.

In the next section we present the asymmetric auction model. The following section presents the non-parametric technique first proposed by Elyakime, et al (1994) and the extension to the asymmetric case as in Laffont and Vuong (1997). Specifications issues are discussed thereafter. Estimation results are then presented.

4.6 Asymmetric private value auction

We can relax the symmetry assumption by postulating a different (continuous) distribution, say $G(c)$ with $c \in [d, d]$, for the costs of DSOs. Then we have an asymmetric private values auction\(^{28}\).

The differences between $G(c)$ and $F(c)$ will reflect natural or artificial cost advantages favouring the DSOs. For example, if local authorities make private firms post an onerous performance bond, which DSO do not have to post, the we would expect costs for private firms to be higher than those of DSOs. Another interpretation would be that CCT provides a fair playing field but that ownership structure, or incumbency advantage, makes one type of firm more efficient than the other.

The probability of the DSO winning a contract is equal to the probability that the costs of all the private firms is above the level that would induce them to outbid the DSO. We define $\sigma(c)$ as the bidding function of the (symmetric) private firms that maps realised costs to a bid. Under fairly general conditions the equilibrium bidding functions will be strictly increasing and differentiable (Bajari (1997)). If $\sigma_p^{-1}(b)$ denotes the inverse of the bidding function used by private firms, the probability of the DSO outbidding a private firm is $(1 - F(\sigma_p^{-1}(b)))$, where $b$ is the bid of the DSO. Assuming that firms costs

\(^{28}\)The $n - 1$ private firms are still assumed to be symmetric within that class.
are independent, the expected utility of a DSO with cost $c$ would be,

$$U_{dso} = (b - c) \left( 1 - F(\sigma_p^{-1}(b)) \right)^{n-1}. \quad (4.10)$$

Applying the same logic to a private firm gives the following expected utility,

$$U_p = (b - c) \left( 1 - G(\sigma_{dso}^{-1}(b)) \right) (1 - F(\sigma_p^{-1}(b)))^{n-2} \quad (4.11)$$

where $\sigma_{dso}$ is the bidding strategy used by the DSO. To find the Bayesian Nash equilibrium strategies of this auction we maximise (4.10) and (4.11) with respect to $b$ and solve for the equilibrium bid functions from the resulting first order conditions.

Unlike the symmetric case, the equilibrium bidding functions cannot be solved explicitly. Instead, the first order conditions of the equilibrium is described by the following pair of differential equations:

$$\frac{\partial \sigma_p}{\partial c}(\sigma_p^{-1}(b)) = (b - c)(n - 1) \frac{f(\sigma_p^{-1}(b))}{1 - F(\sigma_p^{-1}(b))} \quad (4.12)$$

$$\frac{\partial \sigma_{dso}}{\partial c}(\sigma_{dso}^{-1}(b)) = \left( b - c \right) \frac{g(\sigma_{dso}^{-1}(b))}{1 - G(\sigma_{dso}^{-1}(b))} + (n - 2)(b - c) \frac{f(\sigma_p^{-1}(b))}{1 - F(\sigma_p^{-1}(b))} \frac{\partial \sigma_{dso}}{\partial c}. \quad (4.13)$$

The boundary condition for the above system is $\sigma_p(\bar{c}) = \sigma_{dso}(\bar{c}) = \bar{c}$. Also, for an Bayesian Nash equilibrium strategy to exist, the optimal bidding functions must be such that $\sigma_p(\underline{c}) = \sigma_{dso}(\underline{c}) = b^*$ for some $b^* \in (\underline{c}, \bar{c})$.

---

4.7 Estimation method

The object of structural estimation is to recover the distributions $F(c)$ and $G(c)$ using observed bid data from a series of auctions. Several approaches have been proposed in the literature$^{30}$. The structural econometrics of auction literature was reviewed in chapter 1. Here we present a brief summary.

One alternative is to make a parametric assumption about the cost distributions and estimate the parameters of these distributions using maximum likelihood. Unfortunately, the likelihood approach has several drawbacks. In the first place, to implement this approach the equilibrium bidding functions, $\sigma_p(c)$ and $\sigma_{ds}(c)$, must be inverted. In the case of the asymmetric private values auction these functions are not solvable analytically, rather they are expressed as the differential equation system (4.12) and (4.13). Estimation by maximum likelihood would require solving numerically the above system for each auction at each trial parameter vector. Although feasible in principle, it would require large amounts of computer time$^{31}$.

In addition, the maximum likelihood approach suffers from another limitation. In a first price auction, the lowest possible bid, $b = \sigma(c)$, will depend on the parameters of the distribution functions since the equilibrium bidding strategies depend on $G$ and $F$. Thus, the range of possible values of bids will depend on the parameters of interest, which violates one of the regularity conditions needed to prove consistency and asymptotic normality of the maximum likelihood estimator$^{32}$.


$^{32}$The constrained maximum likelihood estimator proposed by Donald and Paarsch (1996) could be used. This estimator is consistent and its asymptotic properties have been derived for the case of discrete covariates. However, the numerical problems associated with the estimation of an asymmetric auction model would be compounded with
Laffont, Ossard and Vuong (1995) propose an alternative to maximum likelihood to estimate the first price auction model. However, their simulated non-linear least squares estimator (SNLLS) is not suitable for our purposes. The SNLLS relies on the revenue equivalence theorem. For asymmetric auctions the revenue equivalence theorem does not hold. Therefore, it is not possible to use a simple simulator for the first moment of the winning bid.

Both, the maximum likelihood approach and moment estimators, such as the SNLLS, require parametric assumptions regarding the distribution of costs. As was discussed before, the performance of different auction mechanisms in an asymmetric context depends crucially on the relative shapes of the underlying distributions. It is preferable, therefore, not to restrict these distributions to a particular parametric form.

Other estimation methods, such as Deltas (1995) are not applicable to the present situation. The method relies on observing all of the bids not just the winning bid. Unfortunately, in our sample we only observe the winning bid for each contract.

The most appropriate estimation technique for our data is that proposed by Elyakime et al. (1994) and Guerre, et al. (1995). The crucial insight of these authors was to substitute the distribution of bids into the differential equation system (4.12) and (4.13) defined above. Defining $D(b)$ as the distribution function of bids for the DSO and $P(b)$ as the distribution function of bids for private firms, we have the following relations,

$$
D(b) = Pr \left( b < b \right) = Pr \left( c < \sigma_{dso}^{-1}(b) \right) = G \left( \sigma_{dso}^{-1}(b) \right) = G(c) \tag{4.14}
$$

$$
P(b) = Pr \left( b < b \right) = Pr \left( c < \sigma_{p}^{-1}(b) \right) = F \left( \sigma_{p}^{-1}(b) \right) = F(c) \tag{4.15}
$$

and

$$
d(b) = \frac{g(c)}{\sigma_{dso}^2} \tag{4.16}
$$

such an estimator.
The last two relations are the densities of bids. These are obtained by differentiating the distributions (4.14) and (4.15) with respect to $b$ and noting that \( \frac{\partial \sigma^{-1}(b)}{\partial b} = \left( \frac{\partial \sigma(c)}{\partial c} \right)^{-1} \). Using these changes of variables in (4.12) and (4.13) we obtain,

\[
c_{dso} = b_{dso} - \frac{1}{n-1} \frac{1 - P(b_{dso})}{p(b_{dso})} \quad (4.18)
\]

\[
c_p = b_p - \frac{1}{d(b_p) + (n-2)\frac{p(b_p)}{1-P(b_p)}}. \quad (4.19)
\]

This ingenious transformation lends itself to a natural non-parametric estimation technique. Guerre, et al. (1995) propose a two stage estimation method. First use the bid data to estimate $d(b), D(b), p(b)$, and $P(b)$ using non-parametric kernel estimators. Then, use the estimated functions, say $\hat{d}(b), \hat{D}(b), \hat{p}(b)$, and $\hat{P}(b)$, and equations (4.18) and (4.19) to generate a sample of pseudo costs, $\hat{c}_p$ and $\hat{c}_{dso}$ for each bid in the data. Finally, use the generated pseudo costs to estimate $G(c)$ and $F(c)$ by kernel estimation techniques.

There is the important issue of identification for the above estimator. The identification question in this context is whether the distributions $F(c)$ and $G(c)$ are uniquely determined from the observed bids and the number of bidders in each auction. Laffont and Vuong (1997) prove that the underlying distributions are in fact identified if and only if the functions defined by (4.12) and (4.13) above are both strictly increasing and differentiable on $[b, \tilde{b}]$.

One final complication arises from the fact that we do not observe all of the bids for a particular contract, only the winning bid and whether the winner was the DSO or a private firm. Therefore, we cannot estimate $D(b)$ and $P(d)$, and the above estimator is not applicable. Laffont and Vuong (1997) extend the above framework to the case where only the winning bid is observed.
Define $D^w(b)$ as the probability distribution of the winning bid conditional on the DSO winning, and $P^w(b)$ probability the winning bid conditional on a private firm winning. These distributions can be estimated from the observed data, since we observe the winning bid and the identity of the winner. Further, define $\pi_{dso}$ as the probability of the DSO winning and $\pi_p$ as the probability of a private firm winning. Following analogous steps as above, conditions (4.12) and (4.13) can be expressed as

\[
_c_{dso} = b_{dso} - \frac{(1 - D^w(b_{dso})) \pi_{dso} + (1 - P^w(b_{dso})) \pi_p}{P^w(b_{dso})\pi_p}
\]

(4.20)

\[
_c_p = b_p - \frac{(1 - D^w(b_p)) \pi_{dso} + (1 - P^w(b_p)) \pi_p}{d^w(b_p)\pi_{dso} + \frac{n-2}{n-1}P^w(b_p)\pi_p}
\]

(4.21)

Again, Laffont and Vuong (1997) show that the distribution functions $G(c)$ and $F(c)$ are uniquely identified provided that the functions defined by (4.20) and (4.21) are strictly increasing and differentiable on the range of bids.

The functions defined above suggest a straightforward estimation strategy. First use the bid data and the identity of the winner to estimate $G^w, F^w, g^w$ and $f^w$, using some non-parametric technique. Estimate $\pi_p$ and $\pi_{dso}$ by the observed success rate of each type of company in the sample. Use the estimated functions, $\hat{G}^w, \hat{F}^w, \hat{g}^w, \hat{f}^w, \hat{\pi}_p$ and $\hat{\pi}_{dso}$, to generate a pseudo-sample of costs using equations (4.20) and (4.21) and each of the winning bids in the sample.

The estimated costs are the values associated with each winning bid, and, therefore, constitute draws from the conditional (on winning) distribution of cost\(^{33}\). Comparing these costs between the DSO and private firms will indicate if there is a cost advantage to one type of firm or the other.

\(^{33}\)Laffont and Young (1997) show that the whole distribution of costs for both types of firms can be recovered, not just the distribution of costs conditional on winning. However we do not undertake this additional step.
4.8 Specification, estimation and results

The data set used was already discussed in section 4.3 above. Because of the small sample size we must make some simplifying assumption. First, to reduce the number of covariates we assume that there are no economies of scale. As the preliminary analysis showed, the hypothesis of constant returns to scale could not be rejected in the reduced form regression models. Second, we assume that there are no other conditioning variables that affect bids besides the number of domestic units (and the number of bidders). Again, the results of the reduced form analysis shows that neither the wage rate, non-domestic units served, nor other explanatory variables seem to be very important in determining costs. The above two assumptions allow us to compare bids across local authorities using a single variable, the average net expenditure per domestic unit served. This variable will be taken as the relevant bid of firms.

Another crucial assumption is that, in the following analysis, the number of bids received for a contract is assumed to be exogenous and known by all bidders. A shown in the preliminary data analysis, there is evidence of a local authority individual effect correlated with costs and with the number of tenders. We ignore those results in what follows.

We first estimate $d^w(b)$, $D^w(b)$, $p^w(b)$, and $P^w(b)$ non-parametrically. Although we assume no other conditioning variables, the distribution and density of bids will be conditional on the number of bidders, $n_i$. We use standard Kernel estimators for these functions. In particular, for the DSOs

$$
D^w(b; n) = \frac{\sum_{i=1}^{I} 1[b_i < b; dso_i = 1]K\left(\frac{n-n_i}{h_n}\right)}{\sum_{i=1}^{I} K\left(\frac{n-n_i}{h_n}\right)}
$$

(4.22)

The discreteness of $n_i$ does not present a problem for kernel estimation (see Bierens (1987)).
\[ \hat{d}_{\text{so}}(b; n) = \frac{1}{h_b} \sum_{i=1}^{I} 1[d_{\text{so}}i = 1 | K(\frac{b-h_{\text{so}}}{h_b})K(\frac{n-n_i}{h_n})] }{ \sum_{i=1}^{I} K(\frac{n-n_i}{h_n}) } \]  

(4.23)

where \( d_{\text{so}}i \) is a dummy variable that equals one if the DSO was awarded contract \( i \), \( 1[\cdot] \) is an indicator function, \( K[u] \) is a suitable kernel, and \( h_b \) and \( h_n \) are bandwidth parameters. The equivalent estimators for the distribution of bids of private firms are

\[ \hat{d}_{\text{p}}(b; n) = \frac{1}{h_b} \sum_{i=1}^{I} 1[b < b; d_{\text{so}}i = 0 | K(\frac{n-n_i}{h_n})} \]  

(4.24)

\[ \hat{p}_{\text{so}}(b; n) = \frac{1}{h_b} \sum_{i=1}^{I} 1[d_{\text{so}}i = 0 | K(\frac{b-h_{\text{so}}}{h_b})K(\frac{n-n_i}{h_n})] }{ \sum_{i=1}^{I} K(\frac{n-n_i}{h_n}) } \]  

(4.25)

Notice that the above estimators give positive weights to bids from auctions with different number of bidders. With a discrete variable it would be more convenient to estimate the distributions using only the bids observed for each particular value of \( n \). However, the size of our sample precludes such a procedure. In some bid cells there are very few observations. For this same reason, \( \hat{\pi}_{p} \) and \( \hat{\pi}_{d_{\text{so}}} \) were estimated using the overall DSO success rate and one minus this success rate.

Also notice that we estimate the density and distribution function separately. Another alternative would be to estimate the only the density and integrate it to obtain the distribution function. However, for technical reasons Guerre, Perrigne and Vuong (1995a) advice against the latter option.

There is one additional point regarding costs. There were 16 contracts for which there was only one bidder. The information on these contracts could not be used in the above analysis since the equilibrium bid functions are only defined for \( n > 1 \). However, a sole bidder should bid \( \bar{c} \), the upper bound of
the cost distribution. No use was made of this information in the estimation stage, but later these bids will be compared with the upper bound of the bidding function estimated below. The final data set contained information on 89 contracts awarded to a DSO and 70 awarded to a private company.

The normal density was used as a kernel. Adaptive bandwidths were used (Härdle (1990)). The estimation proceeded by taking equidistant points in the domain of both functions (taken to be the observed maximum and minimum of the pooled bids) and using the kernel estimates described above. Using these estimated functions, hazard rates were constructed and the pseudo samples were generated according to (4.20) and (4.21).

Unfortunately, we encountered some numerical problems due to the small sample size of the data set. The denominator in equations (4.20) and (4.21) are densities. Over some ranges of bids these densities were very small and close to zero. This occurred because of sparse observations for some bid
ranges, especially when there were few bidders for the contract. For example, since we only observe the winning bid, the high DSO success rate when there were very few bidders (say 2 or 3) means that there are very few observations for private firms in these cases. Consequently, \( p^{\text{br}}(b_{\text{ds}}) \) will be very small and the second term of equation (4.21) will be larger than the bid. The result is that for some observations the cost estimates are negative. Although, this problem is more acute for the DSO cost estimates, there were also some negative cost estimates for private firms with low bid levels.

In order to proceed we adopt the ad-hoc solution of eliminating negative cost observations in the results shown below. There were 26 negative costs for DSOs and 11 for private firms. These costs estimates were omitted from the analysis below.

The numerical problems mentioned above and the solution adopted imply that extreme care must be taken when interpreting the results. The analysis can be considered more an illustration of what can be done if more data were available rather than a solid empirical result. In the future, if the Local Government Management Board were to make public their data on refuse collection contracts, a better analysis can be undertaken. In that case, bid information on all bidders, not just the winning bid, could be used. The sample size would also be considerably larger and richer, permitting the use of other conditioning variables in the cost distributions of firms, such as political control or certain characteristics of the contracts.

Graph 4.1 shows the pseudo costs and their related bids for both types of firms. In general, the bid functions are increasing in the underlying cost of firms, although the relationship is less clear for the DSOs than for private firms. The monotonicity property is necessary and sufficient for identification of the costs distributions.

Graph 4.2 and 4.3 present the non-parametric estimates of the distribution and density functions of costs. Note that, although the distribution of bids is conditioned on the number of bids, the distribution of underlying
costs is not. Table 4.11 presents some summary statistics of these samples.

It is interesting to note that the average cost of the DSOs is smaller than for private firms. The next section presents some tests to see if this difference, as well differences in other moments of the distribution, is statistically significant.

We can now compare the estimated bid functions with the information from contracts were there was only one bidder. As stated above, an agent who knows that he does not face competition from other bidders should submit a tender equal to the upper bound on the cost distribution. Therefore, the bids in a sole bidder contest should be in the upper bound of the cost valuation estimated above. The average tender submitted by sole DSO bidders was £35.84, which is somewhat above the maximum valuation of £30.00 shown on table 4.11. However, the bids ranged from a minimum of £26.44 to a maximum of £69.06. There was also one local authority that received
Figure 4.3: Distribution and density of costs for private firms.

Table 4.11: Summary statistics of cost pseudo-samples

<table>
<thead>
<tr>
<th></th>
<th>Costs by firm type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DSO</td>
</tr>
<tr>
<td>Mean</td>
<td>13.00</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>6.10</td>
</tr>
<tr>
<td>Skewness coef.</td>
<td>0.63</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.04</td>
</tr>
<tr>
<td>Maximum</td>
<td>30.00</td>
</tr>
<tr>
<td>Number of obs.</td>
<td>63</td>
</tr>
</tbody>
</table>
one tender from a private company. Its value was £27.85, well below the maximum of the distribution of costs for private firms.

The information on the contracts with one bidder is not very consistent with the estimation results. A rationalisation for this anomaly is that perhaps sole bidders did face some "virtual" competition in the form of threats to tender by other firms. Generally, when a local authority tenders a contract they first select a list of firms which are then invited to tender. The bidders that finally do submit a bid is a subset of this list. Perhaps firms consider the number of firms invited to tender as a measure of the competition faced, rather than our ex-post measure of the actual number submitted. Although plausible, this casts doubts on the usefulness of the number of bids variable used above.

4.8.1 Comparing the estimated cost distributions

In this section we present several tests to examine the differences between distributions. These test treat the pseudo-samples of costs as random samples\textsuperscript{35}. Also, we again stress that due to numerical problems associated with the low number of observations, results using the generated pseudo samples must be treated with caution.

As noted above, the mean costs are different for each type of firm. The first test in table 4.12 shows that this differences is statistically significant at a 1% confidence level. This is the standard test for equality of means. If both samples are random, then

\[ Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

is asymptotically distributed as a standard normal. This is the statistic shown in table 4.12, calculated by replacing the population variances with

\textsuperscript{35}This is not strictly true for the estimations using bid samples of finite size. Cost estimates for different observations will not be independent because they were generated using estimated functions of the original data.
Table 4.12: Test results for distribution differences

<table>
<thead>
<tr>
<th>Test type</th>
<th>Test results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
</tr>
<tr>
<td>Differences in Mean</td>
<td>-2.6926</td>
</tr>
<tr>
<td>$Z \sim N(0, 1)$</td>
<td></td>
</tr>
<tr>
<td>Differences in Variance</td>
<td>1.8427</td>
</tr>
<tr>
<td>$S \sim F(n_1 - 1, n_2 - 1)$</td>
<td></td>
</tr>
<tr>
<td>Differences in distribution</td>
<td>14.07</td>
</tr>
<tr>
<td>$X \sim \chi^2(J - 1)$</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

- For the last test the number of categories was 8 ($J=8$). See the appendix for a description of how these categories were constructed.

sample variances.

The second test performed was on the equality of variances across samples. Under the null of equal variances, and defining sample 1 as the sample with the highest variance,

$$F = \frac{s_1^2}{s_2^2}$$

will have a $F$ distribution with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom\(^{36}\). This test also rejects the hypothesis of equal variances between samples.

We also present a general test for differences in distributions suggested in Anderson (1996). It is based on a generalisation of Pearson's goodness of fit test. First, the range of the random variable is partitioned into $J$ exhaustive categories. Denote $x_1^j$ and $x_1^j$ as the vectors of frequencies of each

\(^{36}\)This is equivalent to the Goldfeld and Quandt (1965) test when there are no regressors.
sample falling into each category. Under the null of common population distributions,

$$X = v'\Omega^{-1}v$$

is asymptotically distributed as $$\chi^2(J - 1)$$, where

$$v = \frac{x_1^j - x_2^j}{n_1 - n_2}$$

and

$$\Omega = \frac{n_1 + n_2}{n_1 n_2} \begin{pmatrix} p_1(1 - p_1) & -p_1 p_2 & \cdots & -p_1 p_J \\ -p_2 p_1 & p_2(1 - p_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -p_J p_1 & \cdots & \cdots & p_J(1 - p_J) \end{pmatrix}$$

where $$p_j$$ is the probability of the random variable falling into category $$j$$. In applications $$p_j$$ is replaced by the proportion of the pooled sample falling into this category.

To maximise the power of the test, it is recommended that each category have an even number of observations and that this number be greater than or equal to 6 observations. The third row of table 4.12 shows the results of the above test. The hypothesis of equal distributions is just rejected at the 5% confidence level, but not rejected at a 1% confidence level.

The above tests were also performed on the cost distributions after each pseudo-sample was trimmed by 10% (5% at the lower end and 5% at the top end of cost valuations). Results are shown in table 4.13.

The results with the trimmed data show that the hypothesis of equal means and equal variances can be rejected at conventional levels. However,

---

37The breakpoints of each category were not equidistant. The categories were defined so that in an ascending order there were at least 5 observations in each cell for each type of firm. Although it is recommended that each category have 6 observations or more, the Omega matrix could not be inverted with this number of observations. Therefore, the categories were constructed with 5 as the minimum number of observations. See the appendix to this chapter for more details.
Table 4.13: Test for distribution differences (trimmed valuations)

| Test type                     | Statistic | Prob[|X|]  |
|-------------------------------|-----------|--------|
| Differences in Mean           | Z ~ N(0,1)| -3.0825| 0.0020 |
| Differences in Variance       | S ~ F(n₁ - 1, n₂ - 1)| 2.2340 | 0.0017 |
| Differences in distribution   | X ~ χ²(J - 1)| 10.7173| 0.0975 |

- For the last test the number of categories was 7 (J=7). See the appendix for a description of how these categories were constructed.
- 5% of each sample of cost valuations was trimmed from the top and bottom.
the general test of equality of distributions does not reject the hypothesis of equal cost distributions.

In summary, there is evidence that the underlying cost distributions of the DSOs and private firms are different. The strongest result seems to be that the average cost of DSOs is smaller than the average costs of private firms.

The above result suggests that the high success rate of the DSOs in the bidding process could be attributed in part to an underlying cost difference between firm types.

Figure 4.4 combines the cumulative distribution functions for both types of firms and shows that the private firm distribution is stochastically dominated by the DSO distribution\(^{38}\).

---

\(^{38}\)We have not performed a formal test for the hypothesis of stochastic dominance. See Anderson (1996) for a stochastic dominance test.
4.9 Conclusions

There are several conclusions we may draw from the analysis of this chapter. First, the introduction of competitive tendering has significantly lowered the expenditure on refuse collection services for local authorities. In addition, these savings are higher when there are more agents bidding for a contract. Actual competition in the form of more bidders does seem to matter. This holds even when due account is made for local authority specific effects which may be correlated with the number of bidders.

The above conclusion implies that any plans to abolish CCT would probably lead to an increase in the expenditure on refuse collection by local authorities. This result is consistent with the new theory of regulation. In the presence of asymmetric information, auctions are a good mechanisms for extracting informational rents from firms. In their absence, a regulator— even when offering the optimal second best contract— must relinquish some resources to superiorly informed companies.

If CCT is maintained, a relevant question is whether the tendering mechanism can be improved. A first issue is the fairness with which contracts are tendered. The high DSO success rate might indicate that at present there might be a bias in favour of the DSOs.

The structural analysis undertaken above indicates that a bias that creates cost differences between firms is not inconsistent with the data. If true, the policy of raising the required rate of return on public assets that DSOs must earn, may reduce DSO success rate. The rate was raised from 5% to 6% in 1994. This change forces DSOs to present a higher tender when bidding for a contract. However, there is not much evidence that the DSO success rate was reduce after the required rate of return was increased. Table 4.14 shows the DSO success rate for all refuse collection contracts tendered under CCT in six month periods.

Finally, it should be pointed out that in many cases the number of bidders
Table 4.14: DSO success rate by semester

<table>
<thead>
<tr>
<th>Period</th>
<th>DSO success (%)</th>
<th>DSO success (no. of contracts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 1990</td>
<td>80.0</td>
<td>56</td>
</tr>
<tr>
<td>January 1991</td>
<td>73.7</td>
<td>42</td>
</tr>
<tr>
<td>August 1991</td>
<td>52.5</td>
<td>21</td>
</tr>
<tr>
<td>January 1992</td>
<td>50.0</td>
<td>11</td>
</tr>
<tr>
<td>August 1992</td>
<td>25.0</td>
<td>1</td>
</tr>
<tr>
<td>January 1993</td>
<td>23.1</td>
<td>3</td>
</tr>
<tr>
<td>August 1993</td>
<td>25.0</td>
<td>2</td>
</tr>
<tr>
<td>January 1994</td>
<td>27.3</td>
<td>3</td>
</tr>
<tr>
<td>August 1994</td>
<td>56.5</td>
<td>26</td>
</tr>
<tr>
<td>January 1995</td>
<td>50.0</td>
<td>22</td>
</tr>
<tr>
<td>August 1995</td>
<td>71.8</td>
<td>28</td>
</tr>
<tr>
<td>January 1996</td>
<td>67.6</td>
<td>23</td>
</tr>
</tbody>
</table>

- Figures are for contracts tendered during each six month period starting in August 1990.
for contracts under CCT is low. The setting of a reservation price should then be recommended as a way to reduce refuse collection costs. A reservation price acts as an “additional bidder” and thus forces more aggressive bidding when the number of bidders is low.
Appendix 4.A

In this appendix we describe how the segmentation of the cost data was achieved in order to perform the test of equality of distributions. Costs are arranged in $J$ mutually exclusive and exhaustive categories. The breakpoints of each category should be chosen so that they contain roughly equal, but at least 6, observations. With our data this recommendation is difficult to apply since for some ranges the data is sparse and costs for each type of firm do not always overlap. In the end, an arbitrary method for creating the required breakpoints was chosen. Starting from the minimum cost value, the upper breakpoint for the first segment was defined as the minimum value so that at least 6 observation were in that segment for both types of firms. The next segments were defined by applying the same principle to increasingly higher valuations. Using 6 as the minimum number of observations per cell resulted in a singular Omega matrix. Therefore, the minimum number of observations per category was lowered to 5. Table 4.15 shows the breakpoints in the cost range and the number of observations in each category for the test reported in table 4.12.
Table 4.15: Breakpoints and observations for categories used in the equality of distributions test

<table>
<thead>
<tr>
<th>Breakpoint</th>
<th>DSO</th>
<th>Private Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.9158</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>9.2879</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>11.8560</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>13.9710</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>15.1294</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>17.7918</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>21.9354</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>39.5822</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>63</strong></td>
<td><strong>59</strong></td>
</tr>
</tbody>
</table>

- The breakpoints shown above are the upper values of each segment.
Part IV

Conclusions
Chapter 5

Final Remarks

The new theory of regulation emphasises the importance of unobservable latent variables for the optimal design of regulatory mechanisms. On the resource allocation front, unobservable variables usually represent taste heterogeneity among individuals or households. Furthermore, this taste heterogeneity coupled with non-linear prices will result in some kind of sorting behaviour by consumers according to their taste parameter. This has non-trivial implications for empirical analysis, as discussed in chapter 2.

On the production side of the regulatory model, the unobservable variables represent influences on costs that are crucial to determine the optimal cost reduction effort that a regulated firm should exert. The unobservable variable in a production context also has econometric implications (Wolak (1994)).

In both the demand and supply side of a natural monopoly market, empirical analysis must recognise the problems associated with unobservables if consistent parameter estimates are to be obtained. However, the need for consistency in parameter estimates is not the only reason for paying close attention to unobservable latent variables. In many cases, knowledge of these latent processes is needed to answer many of the most interesting questions faced by regulators. Issues related to the design of optimal tariffs or the impact of changes in the regulatory environment can only be addressed if the
distribution of unobservables is known. This calls for estimation techniques aimed directly at recovering these underlying distributions. Structural modelling, as opposed to reduce form approaches, are therefore unavoidable in this context.

In all three essays presented above, some type of structural model was specified and estimated. In the first essay, a structural household demand system was estimated using expenditure data for the United Kingdom. The emphasis was on the modelling of household energy demands. The second essay presents a structural model for the Norwegian bus transport industry that attempts to recover information regarding the distribution of the adverse selection cost parameter of firms. The third essay presents a structural econometric auction model which is used to recover the underlying cost distribution of refuse collection firms in England.

Structural models have the disadvantage that they may be prone to mis-specification. As the discussion in chapter 3 reveals, sometimes many arbitrary parametric assumptions must be made in order to identify all the parameters of interest. However, there seems to be no alternative to structural modelling if quantitative answers are to be given to questions regarding changes in the regulatory environment. In addition, structural modelling, by imposing theoretical restrictions, allow more precise estimates of parameters of interest\(^1\).

The essays in this thesis illustrate the range of regulatory questions that can be answered using structural analysis that otherwise could not be addressed. The energy demand model is used to estimate the welfare effects of tariff changes in the domestic gas market. Although, non-structural approaches could be devised to answer these welfare questions, the presence of two-part tariffs for some goods complicates the analysis. Incorporating the effects of non-linear prices in the specification and estimation of the demand

\(^1\)Of course, this holds only if the imposed restrictions are valid for the problem analysed.
system requires more structure to be placed on the model. The results of the exercise presented in chapter 2 show that tariff rebalancing in the United Kingdom gas market will have a differential impact on households. Most households stand to gain from the change, but some will be made worse off. In addition, the losers form the tariff change are in general poorer and more vulnerable households.

The Norwegian bus transport model is used to estimate the potential benefits of introducing bus route tendering. To answer this question, knowledge of the underlying distribution of firms’ adverse selection cost parameter is crucial. Only a structural approach can hope to estimate this distribution. The model presented in chapter 3 is, to the best of our knowledge, the first attempt to estimate this distribution using a purely econometric approach. The results presented in chapter 3 show that the tendering of bus routes would reduce the level of transfers that the government is currently obliged to give to the bus industry. The magnitude of the estimated transfer savings are within the range announced by the Norwegian government.

Finally, in the last essay a structural asymmetric auction model is presented for the tendering of refuse collection contracts by English local authorities. The model allows for the estimation of bidders’ underlying cost using bid data. Knowledge of these costs and their distribution can be used to improve the auction mechanism. In the last essay, the results are used to explore the possible reasons for the high winning rate of one particular type of firm. Another, less structural, analysis carried out in chapter 4 confirms that more bidders for a contract reduces the price that local authorities must pay the winning firm to undertake the service.

The New Theory of Regulation, developed over the last decade and a half, has captured the attention of policy makers. Governments and regulators the world over have the pressing need to answer important questions regarding the structure and organisation of the interface between the State and markets. One important limitation of the theory has been the difficulty of empirical
implementation. The inability to give numerical answers to certain regulatory questions may be a factor that hampers the theory's usefulness to regulators. However, the main idea of this thesis was to illustrate that empirical research is possible within the framework of the New Theory of Regulation.
Bibliography


244


252


253


254


255


258


