Nonparametric Identification of Incentive Regulation Models*

Isabelle Perrigne
Pennsylvania State University

Quang Vuong
Pennsylvania State University

January 2007

* We dedicate this paper to Jean-Jacques Laffont. Section 2 arose from a term paper we wrote for a course he taught in Spring 2003. Preliminary versions were presented at Johns Hopkins University, Penn State University, SITE Meeting, European University Institute in Firenze, Université de Toulouse, Columbia University, New York University and Princeton University. We thank P. Courty, J. Harrington, A. Pakes, and the participants at seminars and conferences for much constructive comments. Financial support from the National Science Foundation under Grant SES–0452154 is gratefully acknowledged.

Abstract

This paper establishes the nonparametric identification of the incentive regulation model with ex post observed cost developed in Laffont and Tirole (1986). We first extend the model to allow for general random demand and cost functions, while considering a monopolist producing a private good. We then map the resulting model into a structural econometric model that includes observed and unobserved heterogeneity. We establish the nonparametric identification of the cost of public funds, the demand, cost and effort functions, as well as the joint distribution of the random elements of the structural model, which are the firm's type, the demand and cost shocks, and the unobserved heterogeneity.

Fields: Regulation, Optimal Contracts, Cost Efficiency, Nonparametric Identification, Unobserved Heterogeneity.

Address: Pennsylvania State University, Department of Economics, University Park, PA 16802.

Nonparametric Identification of Incentive Regulation Models

I. Perrigne & Q. Vuong

1 Introduction

Over the past thirty years, economists have emphasized the fundamental role played by asymmetric information in economic relationships. The imperfect knowledge of key economic variables induces strategic behavior among economic agents. A simple example of information asymmetries is an auction, in which the seller does not know bidders' values for the auctioned object and the bidders do not know their competitors' values. As a response, bidders play strategically. Game theory provides a useful tool for analyzing such behaviors through the Bayesian Nash equilibrium. Contracts provide another important example of how information asymmetry governs relationships between a principal who designs the contract and an agent. Two types of imperfect information can affect contractual relationships, namely some agent's hidden characteristics or type and some agent's hidden action or effort leading to the socalled adverse selection and moral hazard problems, respectively.[1] The agent plays strategically as he can misreport his own characteristics and minimize effort. Thus, the principal has to give to the agent appropriate incentives to alleviate such problems through the terms of the contract. See the book by Laffont and Martimort (2001) on the theory of incentives.

[1]An auction model can be viewed as a model with adverse selection where bidders' values can be considered as hidden characteristics. The parallel between auctions and contracts will be emphasized later in the paper.

Contracts are widely used in the economic world. To name a few, agriculture, insurance, retailing and management provide many examples of contractual relationships. In this paper, we are interested in regulatory contracts between a regulatory authority or regulator and a monopolistic firm. The government regulates firms to prevent natural monopolistic behavior. Public agencies may also be concerned by redistribution aspects. Several reasons motivate our interest. First, regulation governs many industries such as utilities (electricity, water, gas, telecommunications) as well as many services (postal service, public transit, railroad), which represent an important component of the economy. Second, data on regulatory contracts are more readily available than on "private" ones. Given the public nature of regulatory commissions, data are in general accessible to the analyst in contrast to private contracts, in which data may be subject to confidentiality issues. Third, regulatory contracts are in general well defined in terms of the objectives assigned to the regulated firm and the compensation arrangements made by the regulator.[2] Fourth, the economic literature provides a solid background to analyze regulation in a framework of imperfect information as surveyed by Baron (1989), Laffont and Tirole (1993) and Laffont (1994).[3]

The incentive regulation model introduced by Laffont and Tirole (1986) represents a breakthrough in the new economics of regulation. This paper explains the trade-off faced by the regulator between firm's rent extraction and efficiency. This trade-off is the key issue in incentive regulation in presence of information asymmetries. Namely, the regulator has to provide some incentives to the firm to reveal its information or type as well as to exert appropriate effort. Though the regulator would like to extract all the firm's rent as leaving rents to the firm is costly to society, the regulator must give up some rent to achieve revelation and efficiency from the firm. Moreover, subsidizing the firm through some monetary transfer requires additional taxes, which are costly to society.[4] Laffont and Tirole (1986) have shown that the regulator can achieve such a trade-off while using

[2]Private contracts can be under an implicit form or subject to many unspecified terms such as in incomplete contracts.

[3]Laffont (1994) refers to the new economics of regulation in opposition to the classical theory of regulation neglecting information asymmetries. The latter literature has provided, however, important pricing rules such as the Ramsey-Boiteux and the peak-load pricing rules.

[4]The Laffont and Tirole (1986) model considers a case when transfers are legally possible. By federal law, such transfers may be forbidden. Transfers in this case may take a different form.

the expost observable costs in the transfer paid ex post to the firm.[5] This results in an increase of the social welfare relatively to the Baron and Myerson (1982) model in which no expost cost information is used.[6]

In contrast to these important theoretical developments and the economic importance of regulation in society, very few empirical studies relying on a structural modeling have been performed. As a matter of fact, the empirical literature is mostly limited to a reduced form approach as surveyed by Joskow and Rose (1989).[7] Though this literature acknowledges the presence of information asymmetries, very few papers have attempted to estimate a regulatory contract model. Notable exceptions are Wolak (1994) and Brocas, Chan and Perrigne (2006) for the regulation of water utilities, and Gagnepain and Ivaldi (2002), Perrigne (2002) and Perrigne and Surana (2004) for public transit.[8] This clearly did not meet the expectation of theorists. Laffont (1994, p. 532) writes that the paper by Wolak (1994) "is the first in a long series of applied works which will renew the econometrics of regulation with the help of the new theory of regulation." Similarly, Laffont and Tirole (1993, p. 669) conclude that "econometric analyses are badly needed in the area," while they "do wish that such a core of empirical analysis will develop in the years to come."

Such high expectations have not been met because of the complexity of the models to be estimated. Asymmetric information models lead to highly nonlinear models whose estimation requires suitable econometric tools. Moreover, the issue of identification needs to be addressed.[9] Parametric identification can in principle be achieved but the resulting em-

[5]Firms are usually submitted to an annual audit of their financial results and costs by the regulatory commission. In western economies, accounting rules are well defined and such data are reliable. In developing countries, the problem is somewhat different. Morever, the cost of public funds may be large. Both characteristics lead to different incentive rules as shown by Laffont (2005).

[6]In addition to these nice features, the Laffont and Tirole (1986) model combines both adverse selection and moral hazard by reducing the latter to a "false" moral hazard problem.

[7]Empirical studies have analyzed the effects of regulation on price, product quality, innovation and productivity growth to name a few. The determinants of regulatory mechanisms have also been analyzed. Some of these studies rely on natural experiments when a change in the regulatory process takes place.

[8]The situation is quite similar for the analysis of contract data in general. See the survey by Chiappori and Salanié (2003).

[9]Another important related question is to derive the restrictions imposed by the model on observables to test the model validity. Without such restrictions, the model could rationalize any data. The problem

pirical findings are questionable as identification closely depends on particular functional forms.[10] Moreover, important misspecification issues may arise. The recent literature on the structural analysis of auction models constitutes a stepping stone from which the econometrics of contract models can develop.

In this paper, we adopt a nonparametric approach in the spirit of Guerre, Perrigne and Vuong (2000) to address the identification of the incentive regulation model developed by Laffont and Tirole (1986). Our results are general in the sense that other contract models can be identified using a similar approach. We first need to adapt their model. In particular, they consider the regulation of a public good with a fixed demand, while a private good with a random demand seems to be the most prevalent case in regulation.[11] The contract design then needs to consider expected demand, while the firm will have to fullfill the realized demand. Additional difficulties lie in the contract implementation and in checking whether the local second-order conditions are globally satisfied. In this respect, some assumptions need to be revised accordingly.

We derive the corresponding structural econometric model. In particular, the error terms must arise naturally from the theoretical model. We then face a number of complications. The structural elements are a demand function subject to some random shock, a cost function depending on the unobserved firm's type and effort subject to some random shock, the cost of public funds, the effort disutility function and the firms' type distribution. The observables are the (ex post) demand, the (ex post) cost, the price decided by the regulator and the transfer paid to the firm. A first difficulty arises from the fact that the firm's effort and type, which can be viewed as firm's unobserved heterogeneity, are both unobserved. A second difficulty is related to the singularity of the model. Three unobserved random variables (demand and cost random shocks and firm's type) determine four endogenous variables (demand, cost, price and transfer). Thus, the econometric

of deriving restrictions and testing models is addressed in Perrigne and Vuong (2007a).

[10]The econometrics of auction models provides interesting examples in this respect. For instance, the common value auction model can be estimated parametrically using specific distributions for the structure of the model, while this model is not identified in general. See Paarsch (1992) and Février, Preget and Visser (2004) for parametric estimation of common value auction models. See Laffont and Vuong (1996) and Li, Perrigne and Vuong (2000) for the nonidentification of the common value model.

[11]See the previous examples, in which the consumer needs to pay to get access to the regulated good.

model is singular. We thus introduce an additional term representing unobserved heterogeneity, which can be added equivalently to the effort disutility function or the transfer function. An additional advantage of such an error term is that it can be used to assess the explanatory power of the Laffont and Tirole's model. The econometric model is then based on five equations, which are the demand, the cost, the *generalized* Ramsey pricing, the optimal effort level and the transfer.[12]

The econometric model allows for exogenous variables capturing observed firm, regulator, and/or market heterogeneity, which can affect all the functions and distributions in the model. We first show that the model is nonparametrically identified given the cost of public funds under a multiplicative decomposition of the cost function into a base cost function and an inefficiency cost function. The latter gives the firm's cost inefficiency level and is assumed to be the identity function in this basic model. The firm's type is assumed to be conditionally independent of the three other random shocks in the models, while a natural location-scale normalization is imposed.[13] Nonparametric identification then relies on the bijective mapping between the price and the firm's type. Using the price distribution allows us to identify the optimal effort level as well as the firm's type. The analogy with the auction model becomes clear. In particular, the bidder's private value and his bid in an auction model are equivalent to the firm's type and the price in a regulatory contract. Guerre, Perrigne and Vuong (2000) recover the bidder's private value from the equilibrium first-order condition, which relates monotonically the private value to the bid and the bid distribution. A similar idea is exploited here.

Next, we address the nonparametric identification of the cost of public funds, which requires additional assumption(s). We then consider a model in which the cost inefficiency function is of general known form. The identification follows a similar argument. On the other hand, when considering the case of an unknown cost inefficiency function, we show that there exists an observationally equivalent model, where the cost inefficiency function

[12] The pricing rule derived from the model does not correspond exactly to the well known Ramsey pricing rule because of the stochastic demand. We then call it the *generalized* Ramsey pricing. The *generalized* Ramsey pricing and the optimal effort level equations are directly derived from the first-order conditions of the regulator's maximization problem.

[13] Such a normalization is necessary as a linear transformation of the firm's type will lead to an observationally equivalent model causing the nonidentification of the model.

is the identity function. Thus the general model is not identified and without loss of explanatory power the cost inefficiency function can be taken to be the identity function.

Our paper represents an important step towards the development of the econometrics of contract models and more generally of models with incomplete information. First, we show that the Laffont and Tirole (1986) model is nonparametrically identified from observables. The complexity of the model and its technical difficulties raise numerous challenges as noted above. Second, the nonparametric nature of our identification result leads to an estimation method that is robust to misspecification. In contrast, Perrigne (2002) develops a parametric estimation procedure based on a parametric identification of the model.[14] Our paper is also parsimonious in error terms in contrast to Wolak (1994). While considering the estimation of the Baron and Myerson (1982) model, Wolak (1994) proposes a fully parametric estimation method adding measurement error terms to the first-order conditions of the regulator's maximization problem in addition to the demand and cost random shocks and the firm's type. Third, our results offer a new approach to the estimation of cost efficiency that incorporates explicitly the effects of information asymmetry through adverse selection and moral hazard.[15] In particular, Lemma 4 and the following discussion indicate that estimation of a cost frontier as performed in classical production frontier analyses (see, e.g. Gagnepain and Ivaldi (2002)) does not exploit all the information in the theoretical model and may produce a biased estimate of the base cost function. Lastly, given the many ingredients included in the model, our identification result can be extended to other models with incomplete information. For instance, a contract model with adverse selection is a simplified version as the effort disutility and the cost of public funds are not part of the model structure. As such, nonlinear pricing and many contractual data besides regulatory contract data can be analyzed structurally. See

[14]This paper discusses some technical difficulties addressed here such as the generalization of the Laffont and Tirole's model, the necessary normalization of the cost inefficiency index and the conditional independence of the firm's type and the demand shock among others. It deals with three observed endogenous variables (demand, cost, price) determined by three random variables (demand and cost random shocks and firm's type). Thus, the additional term of unobserved heterogeneity is not necessary. In particular, this paper does not use information on the transfer, though it also discusses the need of transfer information to identify possibly semiparametrically the model.

[15]See Park and Simar (1994) and Park, Sickles and Simar (1998) for a survey of the classical literature on production frontier models with recent developments in the semiparametric estimation of these models.

Miravete (2002) for a recent contribution to the empirical analysis of nonlinear pricing.

The paper is organized as follows. Section 2 presents a generalization of the Laffont and Tirole (1986) model with a stochastic demand for a private good as well as its implementation through linear contracts and the verification of the second-order conditions. Section 3 addresses the nonparametric identification of the basic model, in which the cost inefficiency function is the identity function. It includes the derivation of the econometric model, while considering the identification of the structure for a given cost of public funds. Section 4 addresses the nonparametric identification of the cost of public funds under additional assumptions. Section 5 considers the identification of the general model, in which the cost inefficiency function can take any form, known or unknown. Section 6 concludes. Two appendices provide the proofs of our results.

2 The Model

In this section we extend the Laffont and Tirole (1986) model of incentive regulation of a monopolist producing a private good by allowing for general random demand and cost functions.[16] The demand for the private good and the cost for producing it are

$$\begin{aligned} y &= y(p, \epsilon_d) \geq 0 \\ c &= c(y, \theta - e, \epsilon_c) \geq 0, \end{aligned}$$

where y is the quantity of private good, c is the corresponding cost, p is the price per unit of private good, θ represents the firm's (inefficiency) type, e is the level of effort exerted by the firm, and (ϵ_d, ϵ_c) are the demand and cost random shocks, respectively. As usual, θ and e are private information to the firm, where θ is the (scalar) adverse selection parameter known to be distributed as $F(\cdot)$ with density $f(\cdot) > 0$ on its support $[\underline{\theta}, \overline{\theta}]$, $\underline{\theta} < \overline{\theta}$. The random shocks (ϵ_d, ϵ_c) are known to be jointly distributed as $G(\cdot, \cdot)$ *independently* from θ.[17]

[16] This section owes much to Jean-Jacques Laffont's comments.

[17] Note that ϵ_c is assumed independent of θ in Laffont and Tirole (1986), while ϵ_d is void because the demand y is fixed. Given the timing of the contract, this assumption is natural. Laffont and Tirole (1986) also consider a constant marginal cost function, namely $c = (\theta - e)y + \epsilon_c$, while Laffont and Tirole (1993, p.171) consider the cost function $c = H(\theta - e)c_o(y) + \epsilon_c$, which is, except for the additive separability of

The regulator offers a price schedule $p(\tilde{\theta}) \geq 0$ based on the firms' announcement $\tilde{\theta}$ about its true type θ as well as a net transfer $t = t(\tilde{\theta}, c)$ based on $\tilde{\theta}$ and the observed *realized* firm's cost c. The realized cost is paid by the regulator so that t is the net transfer. The random shocks (ϵ_d, ϵ_c) are realized *ex post*, i.e. after contractual arrangements have been made between the regulator and the monopolist. Consequently, the contract is designed *ex ante* based on expected values with respect to (ϵ_d, ϵ_c). Upon accepting the contract, the firm must satisfy the *realized* demand $y = y[p(\tilde{\theta}), \epsilon_d]$ at the price $p(\tilde{\theta})$ corresponding to its announcement $\tilde{\theta}$. The regulator and the firm are both risk neutral.

Throughout, we assume that all functions are at least twice continuously differentiable and that integration and differentiation can be interchanged. Whenever $a(\cdot)$ is a function of more than one variable, we denote its derivative with respect to the kth argument by $a_k(\cdot)$. All discussions of assumptions and second-order conditions are relegated to subsection 2.5.

2.1. The Firm's Problem

Given the price $p(\cdot)$ and transfer $t(\cdot, \cdot)$ functions chosen by the regulator, the realized utility for the firm with type θ when it announces $\tilde{\theta}$ and exerts effort e is

$$U(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) = t\Big(\tilde{\theta}, c(y[p(\tilde{\theta}), \epsilon_d], \theta - e, \epsilon_c)\Big) - \psi(e), \tag{1}$$

where $\psi(e) \geq 0$ is the firm's cost for exerting effort e. Because (ϵ_d, ϵ_c) is *ex ante* unknown and the firm is risk neutral, the firm's optimization problem is

$$(F) \qquad \max_{\tilde{\theta}, e} \mathrm{E}[U(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) \mid \theta] = \int U(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c),$$

where the independence between θ and (ϵ_d, ϵ_c) is used and $\mathrm{E}[\cdot]$ denotes the expectation with respect to (ϵ_d, ϵ_c).

The firm's optimization problem can be solved in two steps. In the first step, the effort level e is chosen optimally given the announcement $\tilde{\theta}$ and the true type θ:

$$(FE) \qquad \max_{e} \mathrm{E}[U(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) \mid \theta] = \int U(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c).$$

This gives $e = e(\tilde{\theta}, \theta)$, which solves the first-order condition (FOC):

$$0 = \mathrm{E}[U_3(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) \mid \theta] = \int U_3(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c), \tag{2}$$

ϵ_c, the cost function we consider starting from Section 2.4.

i.e. using (1), $e = e(\tilde{\theta}, \theta)$ solves

$$\int t_2\Big(\tilde{\theta}, c(y[p(\tilde{\theta}), \epsilon_d], \theta - e, \epsilon_c)\Big) c_2(y[p(\tilde{\theta}), \epsilon_d], \theta - e, \epsilon_c)\ dG(\epsilon_d, \epsilon_c) = -\psi'(e). \tag{3}$$

Denote the corresponding expected utility by

$$U(\tilde{\theta}, \theta) \equiv \mathrm{E}[U(\tilde{\theta}, \theta, e(\tilde{\theta}, \theta), \epsilon_d, \epsilon_c) \mid \theta] = \int U(\tilde{\theta}, \theta, e(\tilde{\theta}, \theta), \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c). \tag{4}$$

In the second step, the firm solves

$$\max_{\tilde{\theta}} U(\tilde{\theta}, \theta)$$

giving $\tilde{\theta} = \tilde{\theta}(\theta)$, which solves the FOC: $U_1(\tilde{\theta}, \theta) = 0$.

2.2. Incentive Constraint

We now consider the Incentive Constraint (IC) arising from the firm telling the truth θ, i.e. $\theta = \tilde{\theta}(\theta)$ for any $\theta \in [\underline{\theta}, \overline{\theta}]$. Thus, we must have $U_1(\theta, \theta) = 0$ for any θ. Equivalently, denoting $U(\theta) \equiv U(\theta, \theta)$ and $e(\theta) \equiv e(\theta, \theta)$, and using $U'(\theta) = U_1(\theta, \theta) + U_2(\theta, \theta)$ and then (4), we obtain

$$\begin{aligned} U'(\theta) &= U_2(\theta, \theta) \\ &= \int U_2(\theta, \theta, e(\theta), \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c) + e_2(\theta, \theta) \int U_3(\theta, \theta, e(\theta), \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c) \\ &= \int U_2(\theta, \theta, e(\theta), \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c) \\ &= \int t_2\Big(\theta, c(y[p(\theta), \epsilon_d], \theta - e, \epsilon_c)\Big) c_2(y[p(\theta), \epsilon_d], \theta - e, \epsilon_c)\ dG(\epsilon_d, \epsilon_c), \end{aligned}$$

where the third equality follows from (2) since $e(\theta) = e(\theta, \theta)$, and the fourth equality follows from (1). Hence, using (3) at $\tilde{\theta} = \theta$ and $e = e(\theta) = e(\theta, \theta)$ gives the incentive constraint

$$U'(\theta) = -\psi'(e), \tag{5}$$

where

$$U(\theta) = \int t\Big(\theta, c(y(p(\theta), \epsilon_d), \theta - e(\theta), \epsilon_c)\Big) dG(\epsilon_d, \epsilon_c) - \psi(e(\theta)) \tag{6}$$

$$e(\theta) = \arg\max_e \int t\Big(\theta, c(y(p(\theta), \epsilon_d), \theta - e, \epsilon_c)\Big) dG(\epsilon_d, \epsilon_c) - \psi(e). \tag{7}$$

2.3. The Regulator's Problem

Not knowing $(\theta, \epsilon_d, \epsilon_c)$, the regulator chooses $[p(\cdot), t(\cdot,\cdot)]$, i.e. the price schedule and the transfer function. Suppose that $[p(\cdot), t(\cdot,\cdot)]$ is such that (i) it is truth telling (so that the incentive constraint (5) is satisfied and the firm exerts the optimal level of effort $e = e(\theta) = e(\theta, \theta)$) and (ii) the monopolist participates for any level of its type θ. Given that the regulated good is private, the *ex post* social welfare when θ is the firm's true type is

$$\begin{aligned} SW(\theta, \epsilon_d, \epsilon_c) &= \int_{p(\theta)}^{\infty} y(v, \epsilon_d)dv + (1+\lambda)\Big\{p(\theta)y(p(\theta), \epsilon_d) \\ &\quad -t(\theta, c(y(p(\theta), \epsilon_d), \theta - e(\theta), \epsilon_c)) - c(y(p(\theta), \epsilon_d), \theta - e(\theta), \epsilon_c)\Big\} \\ &\quad +t(\theta, c(y(p(\theta), \epsilon_d), \theta - e(\theta), \epsilon_c)\Big) - \psi(e(\theta)), \end{aligned}$$

where $\lambda > 0$ is the shadow cost of public funds. Thus, using the independence of θ and (ϵ_d, ϵ_c), the expected social welfare is

$$\begin{aligned} &\int SW(\theta, \epsilon_d, \epsilon_c)dG(\epsilon_d, \epsilon_c)dF(\theta) = \\ &\quad \int_{\underline{\theta}}^{\overline{\theta}} \Big\{ \int \Big[\int_{p(\theta)}^{\infty} y(v, \epsilon_d)dv + (1+\lambda)\Big(p(\theta)y(p(\theta), \epsilon_d) \\ &\quad -\psi(e(\theta)) - c(y(p(\theta), \epsilon_d), \theta - e(\theta), \epsilon_c)\Big)\Big] dG(\epsilon_d, \epsilon_c) - \lambda U(\theta)\Big\} dF(\theta), \end{aligned} \qquad (8)$$

where we have used (6). Therefore, the regulator's optimization problem is

$$(P) \qquad \max_{[p(\cdot), t(\cdot,\cdot), e(\cdot), U(\cdot)]} \int SW(\theta, \epsilon_d, \epsilon_c)dG(\epsilon_d, \epsilon_c)dF(\theta),$$

subject to the incentive and the participation constraints, i.e.

$$U'(\theta) = -\psi'(e(\theta)) \qquad (9)$$

$$U(\theta) \geq 0, \qquad (10)$$

for all $\theta \in [\underline{\theta}, \overline{\theta}]$, where $U(\cdot)$ and $e(\cdot)$ are given by (6) and (7). Note that, without loss of generality, the control functions in the optimization problem (P) include $e(\cdot)$ and $U(\cdot)$ since these functions are determined by $p(\cdot)$ and $t(\cdot,\cdot)$ through (6) and (7). In view of (9), note also that $U'(\theta) < 0$ under the condition that $\psi'(\cdot) > 0$, which is assumed hereafter. Hence, the participation constraint (10) can be written equivalently as $U(\overline{\theta}) \geq 0$ or

$$U(\overline{\theta}) = 0 \qquad (11)$$

because the expected social welfare (8) decreases with $U(\cdot)$.

We now solve this optimization problem. First, we note that the objective function (8) depends on the transfer function $t(\cdot,\cdot)$ only indirectly through $U(\theta)$ and $e(\theta)$, which are given by (6) and (7). This suggests to consider the simpler optimization problem

$$(P') \qquad \max_{[p(\cdot),e(\cdot),U(\cdot)]} \int SW(\theta,\epsilon_d,\epsilon_c)dG(\epsilon_d,\epsilon_c)dF(\theta),$$

subject to (9) and (11) only. In subsection 2.4 on implementation, we will verify that there exists a transfer function $t^*(\cdot,\cdot)$ satisfying (6) and (7) for the solution $[p^*(\cdot),e^*(\cdot),U^*(\cdot)]$ of the optimization problem (P').

We denote the *expected demand at price* p by $\overline{y}(p) \equiv \mathrm{E}[y(p,\epsilon_d)] = \int y(p,\epsilon_d)dG(\epsilon_d)$ where $G(\cdot)$ is the marginal distribution of ϵ_d. Let $\mathrm{E}[a(\epsilon_d,\epsilon_c,\theta)] = \int a(\epsilon_d,\epsilon_c,\theta)\ dG(\epsilon_d,\epsilon_c)$ denote the expectation of a function $a(\cdot,\cdot,\cdot)$ with respect to (ϵ_d,ϵ_c) for fixed θ, or conditional upon θ given the independence of θ and (ϵ_d,ϵ_c). We have

Proposition 1: *The functions $p^*(\cdot)$ and $e^*(\cdot)$ that solve the FOC of the optimization problem (P') satisfy*

$$\frac{p-\widetilde{mc}(p)}{p} = \mu\frac{1}{\tilde{\eta}(p)} \tag{12}$$

$$\psi'(e) = \overline{cs}_e(p) - \mu\frac{F(\theta)}{f(\theta)}\psi''(e), \tag{13}$$

where $p=p^*(\theta)$, $e=e^*(\theta)$, $\mu=\lambda/(1+\lambda)$ and

$$\widetilde{mc}(p) = \frac{\mathrm{E}\Big[c_1(y(p,\epsilon_d),\theta-e,\epsilon_c)y_1(p,\epsilon_d)\Big]}{\mathrm{E}[y_1(p,\epsilon_d)]}$$

$$\tilde{\eta}(p) = -\frac{p\overline{y}'(p)}{\overline{y}(p)}$$

$$\overline{cs}_e(p) = \mathrm{E}\Big[c_2(y(p,\epsilon_d),\theta-e,\epsilon_c)\Big].$$

Note that $\widetilde{mc}(p)$ differs from the expected marginal cost $\overline{mc}(p) \equiv \mathrm{E}[c_1(y(p,\epsilon_d),\theta-e,\epsilon_c)]$ for producing one additional unit to satisfy the random demand $y(p,\epsilon_d)$ at price p. Moreover, $\tilde{\eta}(p)$ is the elasticity of the expected demand $\overline{y}(p)$, which differs from the expected elasticity of demand $\overline{\eta}(p) \equiv \mathrm{E}[-py_1(p,\epsilon_d)/y(p,\epsilon_d)]$. On the other hand, $\overline{cs}_e(p)$ is the expected cost saving for one additional unit of effort at the random demand $y(p, \epsilon_d)$. Thus, (12) can be viewed as a *generalized Ramsey pricing*, while (13) is interpreted as usual with a downward distortion in effort due to the second term arising from asymmetric information. In particular, when $\theta = \underline{\theta}$ so that $F(\theta) = 0$, (13) gives $\psi'(e) = \overline{cs}_e(p)$ and the first-best is achieved for the most "efficient" firm $\underline{\theta}$ as usual. Moreover, when the demand is not random, i.e. ϵ_d has a degenerate distribution so that $y(p, \epsilon_d) = \overline{y}(p)$, we have $\widetilde{mc}(p) = \overline{mc}(p)$ and $\tilde{\eta}(p) = \overline{\eta}(p)$ so that (12) and (13) reduce to the FOC in Laffont and Tirole (1986) with a constant marginal cost function an additive random shock $(\theta - e)y + \epsilon_c$ considered there.

Lastly, the optimization problem (P') is complete by determining the optimal firm's rent $U^*(\theta)$. The latter is obtained by integrating out the incentive constraint (9) subject to the participation constraint (11). This gives

$$U^*(\theta) = \int_\theta^{\overline{\theta}} \psi'[e^*(\beta)] \; d\beta, \tag{14}$$

which is strictly positive whenever $\theta < \overline{\theta}$ since $\psi'(\cdot) > 0$.

2.4. Implementation

Hereafter, we assume that the cost function is multiplicatively separable in $\theta - e$ (see Laffont and Tirole (1993, p.171).

Assumption A1: *The random cost function is of the form*

$$c(y, \theta - e, \epsilon_c) = H(\theta - e) \; c_o(y, \epsilon_c), \tag{15}$$

for some functions $H(\cdot) > 0$ *and* $c_o(\cdot, \cdot) \geq 0$.

The function $c_o(\cdot, \cdot)$ can be viewed as the (random) *base cost* function, while $H(\cdot)$ can be interpreted as the *cost inefficiency function* determining the *cost inefficiency level* $H \equiv H(\theta - e)$ of the firm. Let the *expected base cost* for satisfying the random demand $y(p, \epsilon_d)$ at price p be $\overline{c}_o(p) \equiv E_\epsilon[c_o(y(p, \epsilon_d), \epsilon_c)\Big]$.

Proposition 2: *Given assumption A1, consider the following transfer function*

$$t^*(\tilde{\theta}, c) \;\; = \;\; A(\tilde{\theta}) - \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta} - e^*(\tilde{\theta})]} \left\{ \frac{c}{\overline{c}_o[p^*(\tilde{\theta})]} - H[\tilde{\theta} - e^*(\tilde{\theta})] \right\}, \tag{16}$$

where $e^*(\cdot)$ and $p^*(\cdot)$ are the optimal price and effort functions obtained from (P'), $\tilde{\theta}$ is the firm's announcement, c is the firm's realized cost, and

$$A(\tilde{\theta}) = \psi[e^*(\tilde{\theta})] + \int_{\tilde{\theta}}^{\overline{\theta}} \psi'[e^*(\beta)]\, d\beta. \tag{17}$$

Thus, given the price schedule $p^*(\cdot)$ and the transfer function $t^*(\cdot,\cdot)$, announcing its true type θ and exerting the optimal effort $e^*(\theta)$ satisfy the FOC of the firm's problem (F). Moreover, $[p^*(\cdot), t^*(\cdot,\cdot), e^*(\cdot), U^*(\cdot)]$ solves the FOC of problem (P).

In view of (14), $A(\tilde{\theta}) = \psi[e^*(\tilde{\theta})] + U^*(\tilde{\theta})$. Hence, (16) shows that the transfer is equal to the cost of effort plus the firm's (expected) rent minus a fraction of the cost overrun, where the latter is the discrepancy between the realized cost and the expected cost. In particular, (16) can be viewed as a menu of linear cost-reimbursement rules in realized cost c with slopes and intercepts depending on the firm's announcement $\tilde{\theta}$. Moreover, when $\theta = \underline{\theta}$, (13) and (15) imply that $\psi'[e^*(\underline{\theta})] = H'[\underline{\theta} - e^*(\underline{\theta})]\overline{c}_o[p^*(\underline{\theta})]$ so that the slope coefficient in (16) equals -1 when $\tilde{\theta} = \underline{\theta}$. That is, recalling that t is the net transfer, the most efficient firm, which announces its true type $\underline{\theta}$, chooses a fixed-price contract.

2.5. Second-Order Conditions

Up to now, we have considered only the first-order conditions (FOC). In this subsection we verify that our optimal solution corresponds to a global maximum. In particular, it is fundamental to verify that announcing the true type θ holds not only locally but globally. As usual, we do this ex post by verifying that our solution satisfies the second-order conditions (SOC) for a local maximum, and that these SOC extend globally.

First, we make explicit assumptions on the demand, cost and effort functions. We follow Laffont and Tirole (1986) and adapt their assumptions to the case of stochastic demand and cost functions.[18] Let $V(p, \epsilon_d)$ be the social value of producing the quantity demanded at price p given demand shock ϵ_d

$$V(p, \epsilon_d) = \int_p^\infty y(v, \epsilon_d) dv + (1 + \lambda) p y(p, \epsilon_d),$$

i.e., $V(p, \epsilon_d)$ is the sum of the net consumer surplus and the revenue for the regulator computed at the shadow cost of public funds (see Laffont and Tirole (1993, p.132) when

[18]These assumptions are sufficient but not necessary. In Section 5, we provide necessary and sufficient assumptions for the second-order conditions and implementation to hold.

the good is private). Let the expected social value be

$$\overline{V}(p) \equiv \int V(p,\epsilon_d)dG(\epsilon_d) = \int_p^\infty \overline{y}(v)dv + (1+\lambda)p\overline{y}(p) > 0,$$

where we have used the definition of $\overline{y}(p)$.

Assumption A2: *The demand, cost and effort functions satisfy:*
(i) $\overline{V}'(\cdot) < 0$, $\overline{V}''(\cdot) < 0$,
(ii) $\overline{c}_o(\cdot) > 0$, $\overline{c}_o'(\cdot) < 0$, $\overline{c}_o''(\cdot) \geq 0$,
(iii) $H'(\cdot) > 0$, $H''(\cdot) \geq 0$,
(iv) $\psi'(\cdot) > 0$, $\psi''(\cdot) > 0$, $\psi'''(\cdot) \geq 0$.

Assumption A2-(i) is standard (see Laffont and Tirole (1986, 1993)). Since

$$\begin{aligned} \overline{V}'(p) &= \lambda\overline{y}(p) + (1+\lambda)p\overline{y}'(p) &&= (1+\lambda)\overline{y}(p)\left(\frac{\lambda}{1+\lambda} - \tilde{\eta}(p)\right) \\ \overline{V}''(p) &= (1+2\lambda)\overline{y}'(p) + (1+\lambda)p\overline{y}''(p) &&= (1+\lambda)p\overline{y}'(p)\left(\frac{1+2\lambda}{(1+\lambda)p} + \frac{\overline{y}''(p)}{\overline{y}'(p)}\right), \end{aligned}$$

it follows that assumption A2-(i) is satisfied if the expected demand is not too inelastic, i.e. $\tilde{\eta}(p) > \lambda/(1+\lambda)$, and if the expected demand is not too convex, i.e. $-p\overline{y}''(p)/\overline{y}'(p) < (1+2\lambda)/(1+\lambda)$ when $\overline{y}(p) > 0$ and $\overline{y}'(p) < 0$ as expected. Regarding assumption A2-(ii), the definition of the expected base cost $\overline{c}_o(p) = \mathrm{E}_\epsilon[c_o(y(p,\epsilon_d),\epsilon_c)]$ gives

$$\begin{aligned} \overline{c}_o'(p) &= \mathrm{E}_\epsilon\Big[c_{o,1}(y(p,\epsilon_d),\epsilon_c)y_1(p,\epsilon_d)\Big] \\ \overline{c}_o''(p) &= E_\epsilon\Big[c_{o,11}(y(p,\epsilon_d),\epsilon_c)y_1^2(p,\epsilon_d)\Big] + E_\epsilon\Big[c_{o,1}(y(p,\epsilon_d),\epsilon_c)y_{11}(p,\epsilon_d)\Big]. \end{aligned}$$

Thus, assumption A2-(ii) is satisfied if $c_{o,1}(\cdot,\cdot) > 0$, $c_{o,11}(\cdot,\cdot) \geq 0$, $y_1(\cdot,\cdot) < 0$ and $y_{11}(\cdot,\cdot) \geq 0$, i.e. if the base cost function is strictly increasing and convex in quantity and demand is strictly decreasing and convex in price. The latter conditions are satisfied in general. Assumption A2-(iii,iv) follows Laffont and Tirole (1993, p. 171). In particular, the cost inefficiency level is strictly increasing and convex in $\theta - e$, while the effort cost is strictly increasing and strictly convex in e.

We begin with the firm's optimization problem (F). In particular, for any $(\tilde{\theta},\theta)$ consider the firm's optimization problem (FE) with respect to e.

Lemma 1: *Suppose that the transfer function $t(\cdot,\cdot)$ is weakly decreasing and concave in realized cost c. Given assumptions A1–A2, the effort $e(\tilde{\theta},\theta)$, which solves the FOC (3),* *is uniquely defined and corresponds to a global maximum of the problem (FE). Moreover,* $0 \leq e_2(\theta, \theta) < 1$.

Note that $t^*(\cdot, \cdot)$ is weakly decreasing and concave in realized cost c, as it is linear in c with slope $-\psi'[e^*(\tilde{\theta})]/\{H'[\tilde{\theta} - e^*(\tilde{\theta})]\overline{c}_o[p^*(\tilde{\theta})]\} < 0$. Thus, Lemma 1 applies.

Next, we turn to the incentive constraint (5). The local SOC for $\tilde{\theta} = \theta$ to be a local maximum is $U_{11}(\theta, \theta) \leq 0$, where $U(\tilde{\theta}, \theta)$ is given by (4). As is well known, using the FOC: $U_1(\theta, \theta) = 0$ which must hold for any θ, this SOC is equivalent to $U_{12}(\theta, \theta) \geq 0$. But differentiating (4) and using (1) give

$$U_2(\tilde{\theta}, \theta) = \mathrm{E}_\epsilon\Big[t_2(\cdot)c_2(\cdot)[1 - e_2(\tilde{\theta}, \theta)]\Big] - \psi'[e(\tilde{\theta}, \theta)]e_2(\tilde{\theta}, \theta) = -\psi'[e(\tilde{\theta}, \theta)],$$

where the second equality follows from (3) where $e = e(\tilde{\theta}, \theta)$. Hence

$$U_{12}(\tilde{\theta}, \theta) = -\psi''[e(\tilde{\theta}, \theta)]e_1(\tilde{\theta}, \theta). \tag{18}$$

Because $\psi''(\cdot) > 0$, the local SOC: $U_{12}(\theta, \theta) \geq 0$ is equivalent to $e_1(\theta, \theta) \leq 0$, i.e.

$$e'(\theta) \leq e_2(\theta, \theta), \tag{19}$$

since $e(\theta) = e(\theta, \theta)$ implies $e'(\theta) = e_1(\theta, \theta) + e_2(\theta, \theta)$.

When the transfer function $t(\cdot, \cdot)$ is weakly decreasing and concave in realized cost, as is the case for the linearly decreasing transfer $t^*(\cdot, \cdot)$ given by (16), Lemma 1 implies that a *sufficient* condition for the local SOC (19) to hold is that $e'(\cdot) \leq 0$. The next lemma shows that $e^{*\prime}(\cdot) < 0$ under the following assumption.

Assumption A3: *For any* $\theta \in [\underline{\theta}, \overline{\theta}]$
(i) $\psi''[e^*(\theta)]\overline{V}''[p^*(\theta)] + (1 + \lambda)\Big\{H'[\theta - e^*(\theta)]\overline{c}_o'[p^*(\theta)]\Big\}^2 < 0$
(ii) $[(1 + \lambda)/\lambda]H''[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)]/\psi''[e^*(\theta)] \leq d[F(\theta)/f(\theta)]/d\theta$.

Condition A3-(i) is reminiscent of assumption 1-(iii) in Laffont and Tirole (1986) for the case where $c(y, \theta - e, \epsilon_c) = (\theta - e)y + \epsilon_c$ and y is nonrandom. Condition A3-(ii) is slightly stronger than the usual condition that $F(\cdot)$ is log-concave as in Laffont and Tirole (1993, assumption 1.2). It actually reduces to it when $H(\theta - e) = \theta - e$ so that $H''(\cdot) = 0$.

Lemma 2: *Given assumptions A1–A3 and the transfer* $t^*(\cdot, \cdot)$ *and price* $p^*(\cdot)$ *functions, the local SOC (19) for truth telling is satisfied as* $e^{*\prime}(\cdot) < 0$*. Moreover,* $p^{*\prime}(\cdot) > 0$.

In particular, effort $e^*(\cdot)$ is strictly decreasing in firm's type θ, while price $p^*(\cdot)$ is strictly increasing in firm's type. These agree with the fact that the cost inefficiency level $H^* \equiv H[\theta - e^*(\theta)]$ of the firm is strictly increasing with its type θ because $\theta - e^*(\theta)$ is strictly increasing in θ and $H'(\cdot) > 0$.

It remains to show that $\tilde{\theta} = \theta$ provides a *global* maximum of the firm's utility (4) under the optimal transfer (16). This is accomplished by the next result.

Proposition 3: *Given assumptions A1–A3 and the transfer $t^*(\cdot,\cdot)$ and price $p^*(\cdot)$ functions, truth telling provides the global maximum of the expected utility function $U(\tilde{\theta},\theta)$ given in (4). Moreover, the expected transfer $\bar{t} \equiv E_\epsilon\Big[t^*\Big(\theta, c[y(p^*(\theta),\epsilon_d), \theta - e^*(\theta), \epsilon_c]\Big)\Big]$ for a firm with type θ announcing its true type θ and thus exerting the optimal effort $e^*(\theta)$ is strictly decreasing and convex in the firm's cost inefficiency level H^*.*

In particular, the second part of Proposition 3 ensures that the regulator can use a menu of linear cost-reimbursement rules, as was proposed in subsection 2.4 on implementation. Moreover, because the firm's cost inefficiency level H^* is strictly increasing in firm's type, the expected transfer is strictly decreasing in firm's type, as expected.

3 Identification of the Basic Model Given λ

The structural approach relies on the maintained assumption that the regulator offers the optimal price schedule $p^*(\cdot)$ and optimal transfer function $t^*(\cdot,\cdot)$ to the monopolist who then reveals its true type θ. The incentive regulation model of Section 2 then determines the price $p = p^*(\theta)$ per unit of private good, the effort $e = e^*(\theta)$ exerted by the monopolist, the quantity $y = y(p,\epsilon_d)$ of private good given the realized demand shock ϵ_d, the cost $c = c(y, \theta - e, \epsilon_c)$ for producing y given the realized cost shock ϵ_c, as well as the (net) transfer $t = t^*(\theta, c)$ to the firm. Thus, the structural approach leads to a closely related econometric model explaining (y, c, p, e, t) from the random variables $(\theta, \epsilon_d, \epsilon_c)$.

In this section we detail the specification of the econometric model for the observables taking into account possible observed and unobserved heterogeneity. We then study the identification of the structural elements of the model, which are the demand, base cost, cost efficiency and effort disutility functions, the distribution of the firm's type, the distribution of the demand and cost shocks, as well as the shadow cost of public funds from the distribution of the observables. Throughout, assumptions A1–A3 are maintained.

3.1. The Structural Econometric Model

A number of complications arises. First, the effort exerted by the monopolist is unobserved as are the firm's type θ and the demand and cost shocks (ϵ_d, ϵ_c). Hereafter, we thus assume that only (Y, C, P, T) are observed, where we use capital letters to distinguish random variables from their realizations.

Second, the demand, cost and effort disutility functions may depend on a vector of exogenous variables $Z \in I\!R^d$, where Z includes some characteristics of the firm, regulator and/or market. To allow for such dependencies, the demand, cost and effort disutility functions are defined hereafter as $y(p, z, \epsilon_d)$, $H(\theta - e, z)c_o(y, z, \epsilon_c)$ and $\psi(e, z)$ when $Z = z$. Similarly, the firm's type θ, the demand shock ϵ_d and the cost shock ϵ_c may depend on Z. This is accomplished by introducing the conditional distributions $F(\cdot|z)$ and $G(\cdot, \cdot|z)$ for θ and (ϵ_d, ϵ_c) given $Z = z$. Hereafter, we let $[\underline{\theta}(z), \overline{\theta}(z)]$ denote the support of $F(\cdot|z)$. Moreover, the cost of public funds λ may depend on z, i.e. $\lambda = \lambda(z)$ for some positive function $\lambda(\cdot)$.[19] From such dependencies on z, it follows that the optimal price, transfer and effort functions are of the form $p^*(\cdot, z)$, $t^*(\cdot, \cdot, z)$ and $e^*(\cdot, z)$. The correspondingly revised assumptions A1–A3 are then assumed to hold for every value of Z. Hereafter, we let $\mathcal{Z}$ denote the support of the distribution of Z.

Third, for every value of Z, the four observed endogenous variables (Y, C, P, T) are determined by the three unobserved random variables $(\theta, \epsilon_d, \epsilon_c)$. Thus, the econometric model is *singular*. In particular, the net transfer T is a deterministic function of (P, C, Z). Because $P = p^*(\theta, Z)$, which is strictly increasing in its first argument by Lemma 2, then $T = t^*(\theta, C, Z) = t^*[\theta^*(P, Z), C, Z]$, where $\theta^*(\cdot, Z)$ is the inverse of $p^*(\cdot, Z)$. Hence, if Z is observed together with (Y, C, P, T), in which case Z represents the observed heterogeneity, the structural model will be immediately rejected as soon as the observed values of (C, P, T, Z) do not lie perfectly on the surface $T = t^*[\theta^*(P, Z), C, Z]$. To circumvent such a difficulty, it is necessary to introduce another source of randomness.

[19]Clearly, not all the variables in Z may affect these functions and distributions. For instance, some demand shifters may affect only the demand. The cost function may depend on some firm's specific characteristics, while the cost of public funds may depend on economic activity such as in Perrigne (2002). Though possibly helpful, as shown in Guerre, Perrigne and Vuong (2006) for risk aversion in auction models, exclusion restrictions are not exploited here to achieve nonparametric identification.

There are two simple ways to do so. A first method, which is compatible with the structural approach, is to assume that some heterogeneity entering the effort disutility function but not the demand and cost functions is known to the regulator and the firm but unobserved by the econometrician. This is reasonable as the determinants of the effort disutility function are likely to be less known than the determinants of the demand and cost functions. Assuming that such unobserved heterogeneity in the effort disutility function can be summarized by an additive term ϵ_t so that the effort disutility function is now $\psi(\cdot, z) + \epsilon_t$ instead of $\psi(\cdot, z)$, it can be seen from (16) and (17) that the transfer $T = t^*(\theta, C, Z)$ will include ϵ_t as an additive term. Alternatively, a second method for introducing another source of randomness, which is less structural but retains the content of the theoretical model of Section 2 is to consider that the observed transfer T differs from the optimal transfer $T^* = t^*(\theta, C, Z)$ by an ex post additive random term ϵ_t. Such a random term ϵ_t may arise from measuring T^* with error, as data on transfers are likely to be imprecise. The random term ϵ_t may also represent extra transfers from the regulator to the firm that do not rely on cost efficiency considerations. In particular, the second approach is useful when one believes that the observed transfer is not equal to the optimal transfer. In this case, the Laffont–Tirole (1986) model can be viewed as providing a determinant of the observed transfer, while ϵ_t can then be used to assess deviations from the optimal transfer. Both approaches lead to the same econometric model.

Collecting the preceding remarks, rearranging (12) and (13), where $\widetilde{mc}(P)$ and $\overline{cs}_e(P)$ are replaced by $[H(\theta - e, Z)\overline{c}'_o(P, Z)]/\overline{y}'(P, Z)$ and $H'(\theta - e, Z)\overline{c}_o(P, Z)$, respectively, and combining (16) and (17) evaluated at $\tilde{\theta} = \theta$, the structural econometric model for the endogenous variables (Y, C, P, T) and the unobserved effort e given the exogenous variables Z is defined, under assumptions A1–A3, by the nonlinear nonparametric simultaneous equation model generally implicit in P with nonadditive error terms $(\theta, \epsilon_d, \epsilon_c)$

$$Y = y(P, Z, \epsilon_d) \tag{20}$$

$$C = H(\theta - e, Z)c_o(Y, Z, \epsilon_c) \tag{21}$$

$$P\overline{y}'(P, Z) + \mu\overline{y}(P, Z) = H(\theta - e, Z)\overline{c}'_o(P, Z) \tag{22}$$

$$\psi'(e, Z) + \mu\frac{F(\theta|Z)}{f(\theta|Z)}\psi''(e, Z) = H'(\theta - e, Z)\overline{c}_o(P, Z) \tag{23}$$

$$T = \psi(e, Z) + \int_{\theta}^{\overline{\theta}(Z)} \psi'[e^*(\tilde{\theta}, Z), Z]d\tilde{\theta}$$

$$-\frac{\psi'(e,Z)}{H'(\theta-e,Z)}\left\{\frac{C}{\overline{c}_o(P,Z)}-H(\theta-e,Z)\right\}+\epsilon_t, \quad (24)$$

where $\mu = \mu(Z)$, $P = p^*(\theta, Z)$ and $e = e^*(\theta, Z)$ solve (22)-(23), and a prime denotes derivation with respect to the first argument of a function. Following Section 2, $\overline{y}(p, z)$ and $\overline{c}_o(p, z)$ in (21)–(24) are, conditional upon $Z = z$, the expected demand at price p and the expected base cost for producing the random quantity $y(p, \epsilon_d)$ at price p, i.e.

$$\overline{y}(p,z) = \int y(p,z,\epsilon_d)dG(\epsilon_d|z) \quad (25)$$

$$\overline{c}_o(p,z) = \int c_o[y(p,z,\epsilon_d),z,\epsilon_c]dG(\epsilon_d,\epsilon_c|z). \quad (26)$$

To complete the specification of the econometric model, we make the following assumption on the random elements $(\theta, \epsilon_d, \epsilon_c, \epsilon_t)$.[20]

Assumption B1: *θ is independent of $(\epsilon_d, \epsilon_c, \epsilon_t)$ conditional upon Z with $E[\epsilon_t|Z] = 0$.*[21]

The condition $E[\epsilon_t|Z] = 0$ is a normalization. When ϵ_t arises from some heterogeneity in the effort disutility function that is unobserved by the econometrician but known by the regulator and the firm, then (ϵ_d, ϵ_c) must be independent of θ given (Z, ϵ_t) for the theoretical model of Section 2 to apply. Assumption B1 then holds under the normalization $E[\epsilon_t|Z] = 0$ if ϵ_t and θ are independent conditional upon Z, i.e. if the firm's type does not depend on the unobserved heterogeneity conditional on the observed heterogeneity Z. Alternatively, when ϵ_t is interpreted as a random term directly added to the optimal transfer due to measurement errors and/or extra cost-unrelated transfers, it is reasonable to assume that ϵ_t is independent of θ conditional upon $(Z, \epsilon_d, \epsilon_c)$. Because (ϵ_d, ϵ_c) is independent of θ conditional upon Z for the theoretical model of Section 2 to apply, it follows that θ is independent of $(\epsilon_d, \epsilon_c, \epsilon_t)$ conditional upon Z and hence that assumption B1 is satisfied under the normalization $E[\epsilon_t|Z] = 0$.

To summarize, the observables are (Y, C, P, T, Z), where the endogenous variables (Y, C, P, T) are determined by (20)–(24), while $(e, \theta, \epsilon_d, \epsilon_c, \epsilon_t)$ are unobserved. The struc-

[20]While estimating the Baron and Myerson (1982) model, Wolak (1994) introduces socalled additive measurement error terms in the FOC of the regulator's maximization problem, i.e. in (22) and (23). In contrast, our error terms arise naturally from the model following the structural approach.

[21]A weaker requirement than assumption B1 is that θ is independent of (ϵ_d, ϵ_c) conditional upon Z, and $E[\epsilon_t|\theta, Z] = 0$ for every value z of Z. Only the condition $E[\epsilon_t|\theta, Z] = 0$ is used in establishing (B.6) in Appendix B. For simplicity, we consider assumption B1.

tural elements of the model are the cost-of-public-funds function $\lambda(\cdot)$, the demand function $y(\cdot,\cdot,\cdot)$, the base cost function $c_o(\cdot,\cdot,\cdot)$ with its cost inefficiency function $H(\cdot,\cdot)$, the effort disutility function $\psi(\cdot,\cdot)$, the conditional type distribution $F(\cdot|\cdot)$ given Z, and the joint distribution $G(\cdot,\cdot,\cdot|\cdot)$ of the random terms $(\epsilon_d, \epsilon_c, \epsilon_t)$ conditional upon Z. In short, the structure of the model is given by the vector of seven functions $[y, c_o, H, \psi, F, G, \lambda]$. The identification problem is to assess whether these structural elements can be recovered uniquely from the conditional distribution of (Y, C, P, T) given Z. For definitions of identification in parametric and nonparametric contexts, see e.g. Koopmans (1949), Roehrig (1988) and Prakasa Rao (1992). In subsections 3.2–3.4 we study such a problem for the simpler case where the cost inefficiency function $H(\cdot, z)$ is the identity function so that $H(\theta - e, z)$ and $H'(\theta - e, z)$ are replaced by $(\theta - e)$ and 1, respectively, in (21)–(24). The resulting model is called the *basic model*. The more general model with $H(\cdot)$ either known or unknown is then studied in Section 5.

3.2. Identification of $\psi(\cdot,\cdot)$ and $F(\cdot|\cdot)$ Given $\lambda(\cdot)$

In this subsection we consider the basic model and study the nonparametric identification of the effort disutility function $\psi(\cdot,\cdot)$ and the conditional distribution of firm's type $F(\cdot|\cdot)$ assuming that the public funds cost function $\lambda(\cdot)$ is known. Identification of $\lambda(\cdot)$ is addressed in Section 4.

To begin, a *location-scale normalization* is necessary despite the restriction that the cost inefficiency function $H(\cdot,\cdot)$ is the identity function in the basic model. Intuitively, this arises because θ, the base cost function $c_0(\cdot,\cdot,\cdot)$ and the effort disutility function are unknown. The next result formalizes the necessity of such a normalization.

Lemma 3: *Let $\alpha = \alpha(\cdot) \geq 0$ and $\beta = \beta(\cdot) > 0$ be some functions of Z. Consider the two structures $\mathcal{S} \equiv [y, c_o, \psi, F, G, \lambda]$ and $\tilde{\mathcal{S}} \equiv [\tilde{y}, \tilde{c}_o, \tilde{\psi}, \tilde{F}, \tilde{G}, \tilde{\lambda}]$ in the basic model with assumptions A1–A3 and B1, where $\tilde{y}(\cdot,\cdot,\cdot) = y(\cdot,\cdot,\cdot)$, $\tilde{c}_o(\cdot,\cdot,\cdot) = c_o(\cdot,\cdot,\cdot)/\beta$, $\tilde{\psi}(\cdot,\cdot) = \psi[(\cdot - \alpha)/\beta, \cdot]$, $\tilde{F}(\cdot|\cdot) = F[(\cdot - \alpha)/\beta|\cdot]$, $\tilde{G}(\cdot,\cdot,\cdot|\cdot) = G(\cdot,\cdot,\cdot|\cdot)$ and $\tilde{\lambda}(\cdot) = \lambda(\cdot)$. Thus, the structures $\mathcal{S}$ and $\tilde{\mathcal{S}}$ lead to the same conditional distribution of (Y, C, P, T) given Z, i.e. the structures $\mathcal{S}$ and $\tilde{\mathcal{S}}$ are observationally equivalent.*

As the proof of Lemma 3 indicates, the observational equivalence between $\mathcal{S}$ and $\tilde{\mathcal{S}}$ arises because the unknown firm's type θ can be linearly transformed into a new type $\tilde{\theta} = \alpha(z) + \beta(z)\theta$ for each value z of Z. Several location-scale normalizations can be employed. For instance, one can fix how two quantiles of θ vary with z. A natural choice for these quantiles are $\underline{\theta}(z)$ and $\overline{\theta}(z)$, which correspond to the most and least efficient firms, respectively, when $Z = z$. In this case, a location-scale normalization would be to set $\underline{\theta}(z) = \underline{\theta}_o(z)$ and $\overline{\theta}(z) = \overline{\theta}_o(z)$, where $\underline{\theta}_o(\cdot)$ and $\overline{\theta}_o(\cdot)$ are known functions such as the zero and one functions respectively. Such a normalization, however, is not very convenient as we must have $\theta - e^*(\theta, z) \geq 0$ for all (θ, z) in the basic model to ensure that $C = [\theta - e^*(\theta, Z)]c_o(Y, Z, \epsilon_c) \geq 0$.

A more convenient location-scale normalization, which is used hereafter, is obtained by imposing that the cost inefficiency of the most efficient firm is one and that the optimal effort of the least efficient firm is 0, irrespective of the value of Z. Formally, we impose

Assumption B2: For every value z of Z

$$\underline{\theta}(z) - e^*[\underline{\theta}(z), z] = 1 \qquad \text{and} \qquad e^*[\overline{\theta}(z), z] = 0. \tag{27}$$

Because the optimal effort $e^*(\theta, z)$ is strictly decreasing in θ, which implies that the cost inefficiency $\theta - e^*(\theta, z)$ is strictly increasing in θ, the normalization (27) actually determines $\underline{\theta}(z)$ and $\overline{\theta}(z)$ as in the preceding direct location-scale normalization, though (27) fixes those boundaries endogenously through the optimal effort function $e^*(\cdot, z)$. Moreover, the normalization (27) is much convenient as it ensures that $\theta - e^*(\theta, z) \geq 1$ so that the cost frontier is defined by the most efficient firm, while $e^*(\theta, z) \geq 0$ for all firms, as desired. In other words, from (21) with $H(x, z) = x$, it follows that $c_o(y, z, \epsilon_c)$ can be interpreted as the *cost frontier* for producing y given (z, ϵ_c), while $\theta - e = [\theta - e^*(\theta, z)]/[\underline{\theta}(z) - e^*[\underline{\theta}(z), z]]$ can be viewed as the *relative cost inefficiency* of a firm with type θ relative to the cost efficient firm with type $\underline{\theta}(z)$.

We now turn to the nonparametric identification of the effort disutility function $\psi(\cdot, \cdot)$ and the conditional distribution of type $F(\cdot|\cdot)$. We need a preliminary result, which establishes that the expected demand function $\overline{y}(\cdot, \cdot)$ and the expected base cost $\overline{c}_o(\cdot, \cdot)$ are identified nonparametrically from observations on quantity, price and costs given $\lambda(\cdot)$. Moreover, it is shown that the relative cost inefficiency of the firm can be recovered uniquely from the observables. Let $[\underline{p}(z), \overline{p}(z)]$ denote the support of the conditional distribution $G_{P|Z}(\cdot|\cdot)$ of P given Z.

Lemma 4: *Suppose that $\lambda(\cdot)$ is known in the basic model with assumptions A1–A3 and*

B1-B2. Thus the expected demand function $\overline{y}(\cdot,\cdot)$ and the expected base cost $\overline{c}_o(\cdot,\cdot)$ are uniquely determined by $\underline{p}(\cdot)$ and the conditional means of (Y,C) given (P,Z) as

$$\overline{y}(p,z) \;=\; \mathrm{E}[Y|P\!=\!p,Z\!=\!z] \tag{28}$$

$$\overline{c}_o(p,z) \;=\; \mathrm{E}[C|P\!=\!\underline{p}(z),Z\!=\!z]\exp\left\{\int_{\underline{p}(z)}^{p}\frac{\tilde{p}\overline{y}'(\tilde{p},z)+\mu\overline{y}(\tilde{p},z)}{\mathrm{E}[C|P\!=\!\tilde{p},Z\!=\!z]}d\tilde{p}\right\}, \tag{29}$$

where $\mu=\mu(z)$. Moreover, the relative cost inefficiency is

$$\theta-e^*(\theta,z)=\Delta(p,z)\equiv \mathrm{E}[C|P\!=\!p,Z\!=\!z]/\overline{c}_o(p,z), \tag{30}$$

where $p=p^*(\theta,z)$, and the function $\Delta(\cdot,\cdot)$ satisfies $\Delta(\cdot,\cdot)\geq 1$ and $\partial\Delta(\cdot,\cdot)/\partial p>0$.

It is interesting to note that the expected demand (25) can be obtained by a simple regression of Y on (P,Z) despite the possible correlation between the demand shock ϵ_d and Z in the demand equation (20) under assumption B1, as the latter only ensures that ϵ_d is independent of θ and hence of P given Z.[22] On the other hand, a simple regression of C (or $\log C$) on (P,Z), as used in the estimation of production/cost frontier (see, e.g. Gagnepain and Ivaldi (2002)), does *not* estimate the expected base cost (26).[23] Nevertheless, by exploiting the generalized Ramsey pricing rule (22), Lemma 4 indicates that the expected base cost $\overline{c}_o(\cdot,\cdot)$, which is the expected base cost for the most efficient firm given the normalization (27), can be estimated from (29) by combining appropriately the regressions of Y and C on (P,Z) with the knowledge of $\underline{p}(\cdot)$. Moreover, (30) shows that the relative cost inefficiency $\theta-e^*(\theta,z)$ of a firm can be recovered from the firm's individual values (y,p,c,z) as the function $\Delta(\cdot,\cdot)$ is known from the regression of C given (P,Z) and the expected base cost $\overline{c}_o(\cdot,\cdot)$.

[22] For instance, consider a demand that is additively separable in ϵ_d, i.e. $Y=r(P,Z)+\epsilon_d$. Thus, the regression of Y on (P,Z) recovers the expected demand $\overline{y}(p,z)$ at price p despite the possible correlation between ϵ_d and Z. This is so because $r(p,z)\neq\overline{y}(p,z)=r(p,z)+\mathrm{E}[\epsilon_d|Z]=\mathrm{E}[Y|P=p,Z=z]$.

[23] For instance, consider a typical cost specification of the form $\log C=\log[c_o(Y,Z,\epsilon_c)]+\log(\theta-e)=s(Y,Z)+\epsilon_c+\log(\theta-e)$, where $\log(\theta-e)\geq 0$ in view of (27). The composite error term $\epsilon_c+\log(\theta-e)$ is typically correlated with both $Y=y(P,Z,\epsilon_d)$ and Z under assumption B1. Regression, IV estimation, and ML estimation of this model attempts to estimate the cost frontier $s(y,z)$, which is different from the expected base cost $\overline{c}_o(p,z)$ that is relevant in the FOC (22)-(23) of price and effort. See Perrigne and Vuong (2007b).

Using Lemma 4, the next result establishes the nonparametric identification of the effort disutility function $\psi(\cdot,\cdot)$ and the conditional distribution of type $F(\cdot|\cdot)$ from observations on quantity, price, cost and transfer given $\lambda(\cdot)$. To this end, we use an identification strategy in the spirit of Guerre, Perrigne and Vuong (2000). Specifically, we exploit the bijective mapping between the price P and the firm's type θ from Lemma 2, which shows that $p^{*'}(\theta) > 0$. The parallel with auction models becomes clear. In auction models, the bijective mapping between the bidder's (unobserved) private value and his optimal (observed) bid is used to rewrite the FOC of the bidder's optimization problem in terms of observables. Such an equation expresses the unobserved private value in terms of the corresponding optimal bid, the bid distribution and density, from which one can identify the private value distribution. A similar strategy is used here. In particular, because $\theta^*(P,Z) = \theta$, we are able to replace in (23) the ratio $F(\theta|Z)/f(\theta|Z)$ by $[G_{P|Z}(P|Z)/g_{P|Z}(P|Z)] \times (\partial\theta^*(P,Z)/\partial p)$, where $G_{P|Z}(\cdot|\cdot)$ is the conditional distribution of P given Z and $g_{P|Z}(\cdot|\cdot)$ its corresponding density. Using (23) and the identification of $\psi(\cdot,\cdot)$ from (24), we derive an expression for θ as a function of the observed optimal price, its distribution and density from which we identify the type distribution $F(\cdot|\cdot)$ as shown in the next proposition. In particular, the unobserved firm's type θ can be recovered uniquely from the firm's observed price P and characteristics Z once the various unknown functions have been recovered from data on (Y,C,P,T,Z).

We define the functions

$$\Gamma(p,z) = -\frac{\partial \mathrm{E}[T|P=p, Z=z]/\partial p}{\partial\Delta(p,z)/\partial p} \tag{31}$$

$$R(p,z) = \frac{\mu[G_{P|Z}(p|z)/g_{P|Z}(p|z)] \times \partial\Gamma(p,z)/\partial p \times \partial\Delta(p,z)/\partial p}{\Gamma(p,z) - \overline{c}_0(p,z) + \mu[G_{P|Z}(p|z)/g_{P|Z}(p|z)] \times \partial\Gamma(p,z)/\partial p}, \tag{32}$$

for an arbitrary value (p,z). The functions $\Gamma(\cdot,\cdot)$ and $R(\cdot,\cdot)$ are known from the knowledge of the joint distribution of (Y,C,P,T) conditional upon Z in view of Lemma 4.[24] As seen in the proof, the functions $\Gamma(\cdot,\cdot)$ and $R(\cdot,\cdot)$ exploit the expected optimal transfer from (24) and the FOC for optimal effort (23).

[24] In particular, $\Gamma(\cdot,\cdot)$ can be interpreted as the marginal decrease in the expected transfer T due to a one-unit increase in relative inefficiency Δ. As (33) and (34) below show, $\Gamma(p,z)$ is also the marginal cost of effort, i.e. $\Gamma(p,z) = \psi'(e,z)$, while $R(p,z)$ is the marginal decrease in effort due to a one-unit increase in price, i.e. $R(p,z) = -e'(p,z)$.

Proposition 4: *Suppose that $\lambda(\cdot)$ is known in the basic model with assumptions A1–A3 and B1-B2. Thus the effort disutility function $\psi(\cdot,\cdot)$ is uniquely determined by $\underline{p}(\cdot)$, $\overline{p}(\cdot)$ and the conditional means of (Y,C,T) given (P,Z) as*

$$\psi(e,z) \;=\; \mathrm{E}[T|P\!=\!\overline{p}(z),Z\!=\!z] + \int_0^e \Gamma[p^*(\tilde{e},z),z]\;d\tilde{e}, \tag{33}$$

where $\Gamma(\cdot,\cdot)>0$, $\partial\Gamma(\cdot,\cdot)/\partial p<0$, and $p^*(\cdot,z)$ is the inverse of the optimal effort function $e^*(\cdot,z)$, which satisfies

$$e^*(p,z) = \int_p^{\overline{p}(z)} R(\tilde{p},z)\;d\tilde{p}, \tag{34}$$

with $\partial\Delta(\cdot,\cdot)/\partial p > R(\cdot,\cdot)>0$. Moreover, the conditional means of (Y,C,T) given (P,Z) and the conditional distribution of P given Z uniquely determine the conditional distribution $F(\cdot|z)$ of type given $Z=z$ as the distribution of

$$\theta = \theta^*(P,z) \equiv \Delta(P,z) + \int_P^{\overline{p}(z)} R(\tilde{p},z)\;d\tilde{p}, \tag{35}$$

where P is distributed as $G_{P|Z}(\cdot|z)$, for every value z of Z.

In particular, while the minimal effort $e^*[\overline{\theta}(z),z]=0$ by the normalization (27), (34) implies that the maximal effort (exerted by the efficient firm with type $\underline{\theta}(z)$) is

$$e^*[\underline{\theta}(z),z] = \int_{\underline{p}(z)}^{\overline{p}(z)} R(\tilde{p},z)\;d\tilde{p} > 0. \tag{36}$$

Similarly, the lower and upper bounds of the conditional distribution $F(\cdot|z)$ of type are

$$\underline{\theta}(z) \;=\; 1 + \int_{\underline{p}(z)}^{\overline{p}(z)} R(\tilde{p},z)\;d\tilde{p} > 1 \tag{37}$$

$$\overline{\theta}(z) \;=\; \frac{\mathrm{E}[C|P\!=\!\overline{p}(z),Z\!=\!z]}{\mathrm{E}[C|P\!=\!\underline{p}(z),Z\!=\!z]}\exp\left\{-\int_{\underline{p}(z)}^{\overline{p}(z)} \frac{\tilde{p}\overline{y}'(\tilde{p},z)+\mu\overline{y}(\tilde{p},z)}{\mathrm{E}[C|P\!=\!\tilde{p},Z\!=\!z]}d\tilde{p}\right\} > \underline{\theta}(z), \tag{38}$$

from (35) in view of (29)–(30) and $\Delta[\underline{p}(z),z]=1$ from (27).

The key of Proposition 4 is that the observed price P is in bijection with the unobserved type θ given $Z=z$. Thus, conditioning on (P,Z) is actually conditioning on (θ,Z). That is, (35) can be viewed as *the inverse of the optimal price schedule* $p^*(\theta,z)$. A similar remark applies to the firm's effort e, which can be recovered similarly through (34) from the firm's observed value (p,z).

3.3. Identification of $y(\cdot,\cdot,\cdot)$, $c_o(\cdot,\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot|\cdot)$ Given $\lambda(\cdot)$

Lemma 3 establishes the identification of the expected demand and expected base cost functions $\overline{y}(\cdot, z)$ and $\overline{c}_o(\cdot, z)$ for every price $p \in [\underline{p}(z), \overline{p}(z)]$ and $z \in \mathcal{Z}$. For counterfactual exercises or policy evaluations, one may need to identify the remaining structural elements of the model, which are the demand function $y(\cdot,\cdot,\cdot)$, the base cost function $c_o(\cdot,\cdot,\cdot)$ and the conditional distribution $G(\cdot,\cdot,\cdot|\cdot)$ of $(\epsilon_d, \epsilon_c, \epsilon_t)$ given Z. This is the purpose of the present subsection, which still assumes that the public fund cost function $\lambda(\cdot)$ is known.

Unlike the random term ϵ_t, which enters additively in the transfer equation (24), the demand shock ϵ_d and cost shock ϵ_c do not enter additively in the demand equation (20) and cost equation (21). The problem is reminiscent of Matzkin (2003) who argues that the structural specification of a random demand or a cost function seldom leads to an additive random term as several references given in that paper indicate. When the random term does not enter additively into the relationship between the endogenous variable and the exogenous variables, Matzkin (2003) shows that this relationship is nonidentified nonparametrically and that some normalization is needed to identify nonparametrically the function and the distribution of the error term.[25]

Our problem differs from Matzkin's framework in two aspects. First, Section 3.2 allows us to identify the expected demand $\overline{y}(\cdot,\cdot)$ and the expected base cost $\overline{c}_o(\cdot,\cdot)$. A natural question is whether the knowledge of such functions can help to identify the functions $y(\cdot,\cdot,\cdot)$ and $c_o(\cdot,\cdot,\cdot)$ and the joint distribution of error terms $G_{\epsilon_d,\epsilon_c|Z}(\cdot,\cdot|\cdot)$. Second, we have a simultaneous equation model, which creates a potential endogeneity problem as Y in the base cost function may be correlated with the error term ϵ_c through ϵ_d.[26] While proposing

[25]Several normalizations can be entertained. For instance, consider the demand equation (20), namely $Y = y(P, Z, \epsilon_d)$, where $P = p^*(\theta, Z)$ is independent of ϵ_d given Z in view of assumption B1. Clearly, the model is nonparametrically nonidentified as a monotonic transformation of the demand shock ϵ_d can be compensated by an appropriate transformation of the function $y(\cdot,\cdot,\cdot)$. Thus, an obvious though strong normalization is to impose simply that the distribution of ϵ_d given Z is known and equal to (say) $G^o_{\epsilon_d|Z}(\cdot|\cdot)$. With such a normalization, $y(\cdot,\cdot,\cdot)$ is identified since $G^o_{\epsilon_d|Z}(\cdot|z) = G_{\epsilon_d|P,Z}(\cdot|p,z) = G_{Y|P,Z}[y(p,z,\cdot)|p,z]$ so that $y(p,z,\cdot) = G^{-1}_{Y|P,Z}[G^o_{\epsilon_d|Z}(\cdot|z)|p,z]$. In addition to requiring that the distribution of ϵ_d given Z be chosen, a similar normalization does not seem to be useful to identify the cost equation (21) as Y is not independent of ϵ_c given Z.

[26]For a recent contribution to the nonparametric identification of nonlinear simultaneous equation model with nonadditive error terms, see Matzkin (2005).

a more general normalization than Matzkin's (2003) first specification and showing that this normalization is made without loss of generality, the next lemma establishes that the knowledge of the expected demand and base cost does not help in identifying the desired functions and distributions. We need first to introduce some notations and to make some assumptions following Matzkin's (2003) first specification.[27]

Let $[\underline{\epsilon}_d(z), \overline{\epsilon}_d(z)] \times [\underline{\epsilon}_c(z), \overline{\epsilon}_c(z)]$ be the support of the conditional distribution $G_{\epsilon_d,\epsilon_c|Z}(\cdot,\cdot|z)$ of (ϵ_d, ϵ_c) given $Z = z$. Similarly, let $[\underline{y}(z), \overline{y}(z)]$ and $[\underline{y}(p,z), \overline{y}(p,z)]$ denote the supports of the conditional distributions $G_{Y|Z}(\cdot|z)$ and $G_{Y|P,Z}(\cdot|p,z)$ of $Y = y(P, Z, \epsilon_d)$ given $Z = z$ and $(P, Z) = (p, z)$, respectively. We make the following normalizations, while imposing natural strict monotonocity conditions on the demand and cost shocks (ϵ_d, ϵ_c).

Assumption B3:
(i) For all $z \in \mathcal{Z}$, and all $(\epsilon_d, \epsilon_c) \in [\underline{\epsilon}_d(z), \overline{\epsilon}_d(z)] \times [\underline{\epsilon}_c(z), \overline{\epsilon}_c(z)]$, there exist $p_o(z) \in [\underline{p}(z), \overline{p}(z)]$ and $y_o(z) \in [\underline{y}(z), \overline{y}(z)]$ such that

$$y[p_o(z), z, \epsilon_d] = \epsilon_d \qquad \text{and} \qquad c_o[y_o(z), z, \epsilon_c] = \epsilon_c \tag{39}$$

where $p_o(\cdot)$ and $y_o(\cdot)$ are known.
(ii) The demand and base cost functions $y(p, z, \cdot)$ and $c_o(y, z, \cdot)$ are strictly increasing in ϵ_d and ϵ_c, respectively, for all values (y, p, z), while the conditional distributions $G_{\epsilon_d|Z}(\cdot|\cdot)$ and $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\cdot,\cdot)$ of (ϵ_d, ϵ_c) are nondegenerated and strictly increasing in their first arguments.

Hereafter, C_o is the base cost value, which can be recovered using Lemma 4, namely $C_o = C/(\theta - e^*(\theta, z))$ with $\theta - e^*(\theta, z) = \mathrm{E}[C|P = p, Z = z]/\overline{c}_o(p, z)$.

[27] The second specification in Matzkin (2003) is related to the homogeneity of degree one in z and ϵ, a property that arises frequently in economic theory. See also Matzkin (1994). As a matter of fact, such a restriction seems natural for the base cost function, which should be homogenous in degree one in input prices. This restriction would require to choose some values or functions z_o, ϵ_{co}, C_o and the factor of homogeneity γ satisfying the homogeneity of degree one. It remains unclear how to choose these values as their choice is not without loss of generality. Moreover, because Matzkin's proof relies on the use of $\gamma = \epsilon_c/\epsilon_{co}$, the conditional distribution of ϵ_c can be recovered along a specific value of z, i.e. $(\epsilon_c z)/\epsilon_{co}$. Thus $G_{\epsilon_d,\epsilon_c|z}(\cdot,\cdot|\cdot)$ cannot be identified everywhere. A solution would be to exploit some exclusion restrictions, where the demand depends only on Z_1 and the base cost on $Z = (Z_1, Z_2)$. Moroever, the base cost function would be homogenous of degree one in the variable Z_2 only and the error terms (ϵ_d, ϵ_c) would be independent of Z_2 given Z_1 thereby allowing to recover $G_{\epsilon_d,\epsilon_c|Z_1}(\cdot,\cdot,\cdot)$. However, if Z_2 contains input prices, one can expect some correlation with ϵ_d invalidating this approach.

Lemma 5: *Let $Y = y(P, Z, \epsilon_d)$ and $C_o = c_o(Y, Z, \epsilon_c)$ satisfy assumption B3-(ii) with P conditionally independent of (ϵ_d, ϵ_c) given Z. There exists an observationally equivalent system $Y = \tilde{y}(P, Z, \tilde{\epsilon}_d)$ and $C_o = \tilde{C}_o(Y, Z, \tilde{\epsilon}_c)$ with P conditionally independent of $(\tilde{\epsilon}_d, \tilde{\epsilon}_c)$ given Z satisfying assumption B3-(ii) and*

$$\overline{y}(p, z) = \overline{\tilde{y}}(p, z), \quad \overline{c}_o(y, z) = \overline{\tilde{c}}_o(y, z)$$
$$\tilde{y}(p_o(z), z, \tilde{\epsilon_d}) = \tilde{\epsilon}_d, \quad \tilde{c}_o(y_o(z), z, \tilde{\epsilon}_c) = \tilde{\epsilon}_c$$

for any (p, z) and $(\tilde{\epsilon}_d, \tilde{\epsilon}_c)$, where $p_o(z) \in [\underline{p}(z), \overline{p}(z)]$ and $y_o(z) \in [\underline{y}(z), \overline{y}(z)]$ are arbitrary.

Our normalization is more general than the one in Matzkin (2003) by allowing p_o and y_o to depend on z. If one considers constant values p_o and y_o, as in Matzkin (2003), these values may not satisfy $p_o \in [\underline{p}(z), \overline{p}(z)]$ and $y_o \in [\underline{y}(z), \overline{y}(z)]$ for any $z \in \mathcal{Z}$. Lemma 5 shows that the normalization in assumption B3 does not entail any loss of generality and that $p_o(\cdot)$ and $y_o(\cdot)$ can be chosen arbitrarily. This argument can be easily seen in the demand case when the error term is additive. For instance, let $\tilde{y}(p, z, \tilde{\epsilon}_d) = \tilde{y}(p, z) + \tilde{\epsilon}_d$, where $\tilde{y}(p, z) = y(p, z) - y(p_o(z), z)$ and $\tilde{\epsilon}_d = \epsilon_d + y(p_o(z), z)$ for an arbitrary $p_o(z) \in [\underline{p}(z), \overline{p}(z)]$. Thus, we have $Y = y(p, z, \epsilon_d) = \tilde{y}(p, z, \tilde{\epsilon}_d)$ and $\tilde{y}(p_o(z), z) = 0$ leading to $\tilde{y}(p_o(z), z, \tilde{\epsilon}_d) = \tilde{\epsilon}_d$ thereby satisfying assumption B3. In this sense, the term normalization is appropriate. As a matter of fact, though quite intuitive as the distribution of $y(\cdot, \cdot, \cdot)$ at $p_o(\cdot)$ reduces to that of the error term given $Z = z$, the choice of $p_o(\cdot)$ may be quite puzzling to the analyst. Lemma 5 shows that one should not be concerned by the choice of $p_o(\cdot)$ as long as it satisfies $p_o(z) \in [\underline{p}(Z), \overline{p}(Z)]$. A similar remark applies to $c_o(\cdot, \cdot, \cdot)$ and $y_o(\cdot)$. Moreover, Lemma 5 shows that the knowledge of the expected demand and expected base cost does not help in identifying the demand and base cost functions as well as the joint distribution of error terms.

The next result establishes the nonparametric identification of the demand function $y(\cdot, \cdot, \cdot)$, the base cost function $c_o(\cdot, \cdot, \cdot)$ and the conditional distribution $G(\cdot, \cdot|\cdot)$ of (ϵ_d, ϵ_c) given Z from observations on quantity, price, and costs given $\lambda(\cdot)$. Let $C_o = c_o(Y, Z, \epsilon_c)$ be the (random) *base cost.* Because $C = [\theta - e^*(\theta, Z)]C_o$, where C is observed and $\theta - e^*(\theta, Z) = \Delta(P, Z)$ is identified by Lemma 4, it follows that the base cost C_o can be recovered and its conditional distribution $G_{C_o|Y,P,Z}(\cdot|\cdot, \cdot, \cdot)$ given (Y, P, Z) is identified.

More formally, because $\Delta(P, Z) \geq 1$, we have

$$G_{C_o|Y,P,Z}(c_o|y, p, z) = G_{C|Y,P,Z}[c_o\Delta(p, z)|y, p, z]$$

for any c_o, where $G_{C|Y,P,Z}(\cdot|\cdot,\cdot,\cdot)$ is the conditional distribution of C given (Y, P, Z). This information is used next.

Proposition 5: *Suppose that $\lambda(\cdot)$ is known in the basic model with assumptions A1–A3 and B1–B3.*
(i) The demand function $y(\cdot,\cdot,\cdot)$ and the conditional distribution $G_{\epsilon_d|Z}(\cdot|\cdot)$ of ϵ_d given Z are uniquely determined by the conditional distribution $G_{Y|P,Z}(\cdot|\cdot,\cdot)$ as

$$y(p, z, \epsilon_d) = G^{-1}_{Y|P,Z}\left\{G_{Y|P,Z}[\epsilon_d|p_o(z), z]|p, z\right\} \tag{40}$$

$$G_{\epsilon_d|Z}(\cdot|z) = G_{Y|P,Z}[\,\cdot\,|p_o(z), z]. \tag{41}$$

(ii) Suppose that for all $z \in \mathcal{Z}$, and all $\epsilon_d \in [\underline{\epsilon}_d(z), \overline{\epsilon}_d(z)]$, there exists $p_\dagger(z, \epsilon_d) \in [\underline{p}(z), \overline{p}(z)]$ such that $y_o(z) = y[p_\dagger(z, \epsilon_d), z, \epsilon_d]$. Thus the base cost function $c_o(\cdot,\cdot,\cdot)$ and the conditional distribution $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\cdot,\cdot)$ of ϵ_c given (ϵ_d, Z) are uniquely determined by the conditional distribution $G_{C_o|Y,P,Z}(\cdot|\cdot,\cdot,\cdot)$ as

$$c_o(y, z, \epsilon_c) = G^{-1}_{C_o|Y,P,Z}\left\{G_{C_o|Y,P,Z}[\epsilon_c|y_o(z), p_\dagger(z, \epsilon_d), z]|y, p, z\right\} \tag{42}$$

$$G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z) = G_{C_o|Y,P,Z}[\,\cdot\,|y_o(z), p_\dagger(z, \epsilon_d), z], \tag{43}$$

where $p_\dagger(\cdot,\cdot)$ is identified and $y = y(p, z, \epsilon_d)$.

The proof of (i) follows Matzkin(2003) as θ and hence P are independent of ϵ_d given Z by assumption B1. On the other hand, (θ, ϵ_d) and hence $Y = y(P, Z, \epsilon_d)$ are not independent of ϵ_c given Z because of the endogeneity issue. The proof of (ii) is only slightly more involved as it exploits the additional condition in (ii). This condition says roughly that for any z and any value of the demand shock ϵ_d there exists a price $p_\dagger(z, \epsilon_d)$ for which the output $y[p_\dagger(z, \epsilon_d), z, \epsilon_d]$ is equal to the reference output $y_o(z)$ of assumption B3. Hereafter, it is assumed that such a condition holds. As $y_o(z)$ can be chosen arbitrarily by Lemma 5, this condition actually requires the demand function $y(\cdot,\cdot,\cdot)$ and supports $[\underline{p}(z), \overline{p}(z)]$ and $[\underline{\epsilon}_d(z), \overline{\epsilon}_d(z)]$ to be such that $\bigcap_{(p,\epsilon_d)\in[\underline{p}(z),\overline{p}(z)]\times[\underline{\epsilon}_d(z),\overline{\epsilon}_d(z)]}\{y = y(p, z, \epsilon_d)\}$ is nonempty for every $z \in \mathcal{Z}$.[28]

Lastly, the conditional distribution $G_{\epsilon_t|\epsilon_d,\epsilon_c,Z}(\cdot|\cdot,\cdot,\cdot)$ of ϵ_t given $(\epsilon_d, \epsilon_c, Z)$ is identified from observations on (Y, C, P, T, Z). For, the demand and cost shocks (ϵ_d, ϵ_c) can be recovered from (Y, C, P, Z) through (20)–(21) as $y(\cdot,\cdot,\cdot)$ and $c_o(\cdot,\cdot,\cdot)$ are identified by Proposition 5. The identification of $G_{\epsilon_t|\epsilon_d,\epsilon_c,Z}(\cdot|\cdot,\cdot,\cdot)$ follows immediately from (24), since ϵ_t can be expressed as a function of (C, P, T, Z) and functions that are identified and thus estimable from observations on (Y, C, P, T, Z) by Lemma 4 and Proposition 4. For a simpler expression than (24), see Lemma 5 below. Moreover, because $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\cdot,\cdot)$ and $G_{\epsilon_d|Z}(\cdot|\cdot)$ are identified by Proposition 5, then the joint distribution $G(\cdot,\cdot,\cdot|\cdot)$ of $(\epsilon_d, \epsilon_c, \epsilon_t)$ given Z is identified.

4 The Cost of Public Funds

The previous identification results can be used when the cost of public fund $\lambda(\cdot)$ is known. In the US, $\lambda = 0.3$ is a well accepted value among economists, while the cost of public funds takes larger values in developing countries. On the other hand, in microeconometric studies, one may want to distinguish regulatory contracts according to the regulator and/or market. In this case, identification of the cost of public fund as a function $\lambda(\cdot)$ of some characteristics Z is of interest. This is the purpose of this section. We first show that the cost of public funds is not identified in general. We then propose some identifying conditions for $\lambda(\cdot)$.

4.1. Nonidentification of $\lambda(\cdot)$

To address the nonidentification of $\lambda(\cdot)$, we first extend assumptions A2-A3, which follow Laffont and Tirole (1986). In particular, our assumptions are expressed in terms of the observables (Y, C, P, T, Z) and impose implicit restrictions on $\lambda(\cdot)$. Specifically, we define the error terms as *identified* functions of the observables given $\lambda(\cdot)$. From (20) and assumption B3, the demand error term ϵ_d can be expressed as a function of (Y, P, Z), namely $\phi_d(Y, P, Z)$. Similarly, using (21), (29), (30) and assumption B3, the base cost

[28] It can be relaxed. For instance, following the proof of (ii), one obtains the identification of $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z)$ for those values of (ϵ_d, z) for which there exists a price $p_\dagger(z, \epsilon_d)$ satisfying $y_o(z) = y[p_\dagger(z, \epsilon_d), z, \epsilon_d]$. Similarly, one obtains the identification of $c_o(y, z, \epsilon_c)$ for those values of (y, z), where $y = y(p, z, \epsilon_d)$ and (z, ϵ_d) satisfies $y_o(z) = y[p_\dagger(z, \epsilon_d), z, \epsilon_d]$.

error term ϵ_c can be expressed as a function of (Y, C, P, Z), namely $\phi_c(Y, C, P, Z)$. Using (24), (29), (30), and (33), the transfer error term ϵ_t can be expressed as a function of (Y, C, P, T, Z), namely $\phi_t(Y, C, P, T, Z)$. As in Section 3, the cost efficiency function $H(\cdot)$ is the identity function.[29]

Assumption C1: *The cost of public funds $\lambda(\cdot)$ and the joint distribution of (Y, C, P, T) given Z satisfy*
(i) $\overline{y}(p, z) > 0$, $\mathrm{E}[C|P = p, Z = z] > 0$
(ii) $\Gamma(p, z) > 0$, $\partial\Gamma(p, z)/\partial p < 0$
(iii) $\partial\Delta(p, z)/\partial p > 0$ and $\Gamma(p, z) < \overline{c}_o(p, z)$
for any $p \in [\underline{p}(z), \overline{p}(z)]$ and $z \in \mathcal{Z}$. Moreover,
(iv) for some $p_o(\cdot) \in [\underline{p}(\cdot), \overline{p}(\cdot)]$ and $y_o(\cdot) \in [\underline{y}(\cdot), \overline{y}(\cdot)]$, the random variables $\phi_d(Y, P, Z)$, $\phi_c(Y, C, P, Z)$ and $\phi_t(Y, C, P, T, Z)$ are conditionally independent of P given Z
(v) the conditional distribution $G_{P|Z}(\cdot|\cdot)$ has a strictly positive density on its support $\{(p, z) : p \in [\underline{p}(z), \overline{p}(z)], z \in \mathcal{Z}\}$ with $\underline{p}(\cdot) < \overline{p}(\cdot)$, while the conditional distributions $G_{Y|P,Z}(\cdot|\cdot, \cdot)$ and $G_{C|Y,P,Z}(\cdot|\cdot, \cdot, \cdot)$ are nondegenerated and strictly increasing in their first arguments.

Note that $\mu(\cdot) = \lambda(\cdot)/(1 + \lambda(\cdot))$, while $\Delta(p, z)$ and $\Gamma(p, z)$ are defined in (30) and (31), respectively. Assumption C1-(i) is made implicitly in Section 2 as strictly positive expected demand and cost are generally assumed. As such, these assumptions are quite standard. Assumption C1-(ii) ensures that the expected transfer is strictly decreasing and convex in $\theta - e$ as required by Proposition 3. Regarding assumption C1-(iii), the inequalities $R(p, z) > 0$ and $\partial\Delta(p, z)/\partial p > R(p)$ in Proposition 4 ensure a strictly decreasing effort function and a strictly increasing price schedule, respectively as required by Lemma 2. From (32), $R(p, z) > 0$ is equivalent to have its denominator $\Gamma(p, z) - \overline{c}_o(p, z) + \mu(z)(G_{P|Z}(p|z)/g_{P|Z}(p|z)) \times (\partial\Gamma(p, z)/\partial p) < 0$ since the numerator is negative in view of $\partial\Gamma(p, z)/\partial p < 0$ as required by assumption C1-(ii). This inequality can be rewritten as $\Gamma(p, z) < \overline{c}_o(p, z) - \mu(z)(G_{P|Z}(p|z)/g_{P|Z}(p|z)) \times (\partial\Gamma(p, z)/\partial p)$. On the other hand, $\partial\Delta(p, z)/\partial p > R(p)$ is equivalent to $\Gamma(p, z) < \overline{c}_o(p, z)$ after some

[29] Hereafter, it is assumed that, for any structure in $\mathcal{S}$ defined below, there exists a $p_\dagger(z, \epsilon_d)$ satisfying the additional condition of Proposition 5-(ii). This implies some restrictions which are implicitly included in assumption C1.

elementary algebra. Hence the combination of the two inequalities holds if and only if $\partial\Delta(p,z)/\partial p > 0$ and $\Gamma(p,z) < \overline{c}_o(p,z)$. Assumption C1-(iv) is a direct consequence of assumption B1 combined with $p(\theta, z)$ strictly increasing in θ. Lastly, the first part of assumption C1-(v) follows those on $F(\cdot|\cdot)$, while the second part of assumption C1-(v) parallels the second part of assumption B3-(ii).

For the basic model, we define the set of structures $\mathcal{S} \equiv \{S = [y, c_o, \psi, F, G, \lambda]$: assumptions A1, B1 $-$ B3, hold$\}$. Note that assumptions A2-A3 need not be satisfied by structures in $\mathcal{S}$. A conditional distribution for (Y, C, P, T) given Z is *induced* by a structure $S \in \mathcal{S}$ if it satisfies (20)–(24) for some effort function $e(\theta, Z)$.

Lemma 6: *Let $S \in \mathcal{S}$. If S satisfies assumptions A2-A3, then S induces a conditional distribution for (Y, C, P, T) given Z satisfying assumption C1.*

Lemma 6 shows that assumptions A2-A3 are stronger than assumption C1. This result is not surprising as assumption C1 provide parsimonious conditions for the observables relying on the first-order and second-order conditions and the implementation. As a matter of fact, the next lemma shows that assumption C1 provides necessary and sufficient conditions for the conclusions of Lemmas 1-2 and Proposition 3 to hold, namely (i) the firm's objective function is strictly concave in effort so that there is a unique solution to the firm's effort maximization problem, (ii) the optimal effort is strictly decreasing in θ so that the local second-order condition (19) is satisfied, (iii) the optimal price schedule is strictly increasing in θ, (iv) truth telling provides the global maximum of the firm's problem and (v) the expected transfer is strictly decreasing and convex in the firm's cost inefficiency level. Hence, assumptions A2-A3 are sufficient but not necessary.

Lemma 7: *If $S \in \mathcal{S}$ satisfies the conclusions of Lemmas 1-2 and Proposition 3, then the conditional distribution of (Y, C, P, T) given Z induced by S satisfies assumption C1. Conversely, if the cost of public fund $\lambda(\cdot)$ and the conditional distribution of (Y, C, P, T) given Z satisfy assumption C1, then there exists a structure in S satisfying the conclusions of Lemmas 1-2 and Proposition 3 and rationalizing the observations (Y, C, P, T) given Z.*

Lemma 7 shows that assumption C1 characterizes all the restrictions imposed by the model on observables. The derivation of such restrictions is interesting in the structural approach as it allows the analyst to test the validity of the model. Such an issue is left for future research and is treated in Perrigne and Vuong (2007a).

We are now in a position to address the identification of $\lambda(\cdot)$. The next proposition shows that the cost of public funds is not identified. The proof relies on constructing a structure that is observationally equivalent to the original structure generating the observations (Y, C, P, T).

Proposition 6: *In the basic model consisting of structures S inducing conditional probability distributions for (Y, C, P, T) given Z that satisfy assumption C1, the cost of public funds $\lambda(\cdot)$ is not identified.*

This result is a consequence of Lemma 7. As shown in the proof, it suffices to find another cost of public funds $\tilde{\lambda}(\cdot)$ in a structure $\tilde{S}$ that is observationally equivalent to the original structure S generating the observables (Y, C, P, T) and that satisfies assumption C1. Given that the cost of public funds is not identified, we consider some identifying conditions or assumptions in the next subsection.[30]

4.2. Identifying Conditions for $\lambda(\cdot)$

We need first a lemma that expresses the firm's rent and the realized transfer directly from observations on (Y, C, P, T, Z) and identified functions.

Lemma 8: *In the basic model with assumptions A1, B1-B2 and C1, the firm's transfer can be written as*

$$T \;=\; \mathrm{E}[T|P,Z] - \frac{\partial \mathrm{E}[T|P,Z]/\partial p}{\partial \mathrm{E}[C|P,Z]/\partial p - [P\overline{y}'(P,Z) + \mu \overline{y}(P,Z)]}\Big(C - \mathrm{E}[C|P,Z]\Big) + \epsilon_t \quad (44)$$

where $\mathrm{E}[\epsilon_t|P,Z] = 0$. *Moreover, the firm's expected rent is*

$$U^*(\theta) = \mathrm{E}[T|P\!=\!p, Z\!=\!z] - \mathrm{E}[T|P\!=\!\overline{p}(z), Z\!=\!z] - \int_p^{\overline{p}(z)} \Gamma(\tilde{p}, z) R(\tilde{p}, z) d\tilde{p} \;\geq\; 0 \qquad (45)$$

where $\theta = \theta^*(p, z)$ *as given by (35).*

Equation (44) is interesting for several reasons. First, note that the fraction in (44) is strictly positive as it is equal to $\psi'(e, Z)/\overline{c}_o(P, Z)$. Thus, the ex post transfer T is equal to its expectation $\mathrm{E}[T|P, Z]$ plus a combined residual ν_t, which is the difference between the random term ϵ_t and a positive fraction of the overrun cost $C - \mathrm{E}[C|P, Z]$. Moreover,

[30] Another identification strategy would be to consider the implicit and explicit restrictions on $\lambda(\cdot)$ embodied in assumption C1 and to derive some bounds for $\lambda(\cdot)$ given the distribution of (Y, C, P, T, Z) in the spirit of Manski and Tamer (2002) and Chernozhukov, Hong and Tamer (2006).

because the firm's ex post rent is equal to ex post transfer minus the disutility of effort $\psi(e,z)$, while the expected rent is equal to the expected transfer minus $\psi(e,z)$, it follows that the ex post rent is equal to (45) plus the combined residual ν_t. Second, as indicated in subsection 3.3, when $\lambda(\cdot)$ and hence $\mu(\cdot)$ are known, (44) can be used to recover ϵ_t and hence to establish the identification of $G_{\epsilon_t|\epsilon_d,\epsilon_c,Z}(\cdot|\cdot,\cdot,\cdot)$. Third, (44) suggests how $\mu(\cdot)$ and hence $\lambda(\cdot)$ can be identified.

Equation (44) gives us an expression of the transfer in terms of the observables (Y,C,P,T,Z), the cost of public funds $\mu(\cdot)$ and the unobserved heterogeneity ϵ_t. This suggests that identifying assumptions need to be made on ϵ_t. We note that $\mathrm{E}(\epsilon_t|Z)=0$ or $\mathrm{E}(\epsilon_t|P,Z)=0$ is not sufficient to identify $\mu(\cdot)$ because (44) would lead to an identity. We could then exploit other conditional independence of ϵ_t to drop the term of unobserved heterogeneity in (44). Several assumptions can be entertained. For instance, we could think of an instrumental variable approach and find some instruments Z_1 in the vector of exogenous variables $Z=(Z_1,Z_2)$ such that ϵ_t is independent of Z_1 conditional on Z_2 and θ. More generally, we could find some additional instruments W that are conditionally independent of ϵ_t given (Z,θ). This would lead to the conditional expectation of $W(T-\mathrm{E}[T|P,Z])$ given (P,Z). From (44), the latter would be equal to (say) $K(\mu,P,Z)\mathrm{E}[W(C-\mathrm{E}[C|P,Z])|P,Z]$, since $\mathrm{E}[W\epsilon_t|P,Z]=0$ by assumption. From such an equality, one can recover $\mu(\cdot)$. Though this approach is standard in econometrics, the choice of instruments requires some attention. Morever, the analyst may not have additional exogenous variables beyond those embodied in Z.

A second possibility is to view (44) with a combined error term ν_t as discussed above. In particular, (44) expresses the transfer as the sum of the expected transfer and an error term ν_t, where the first part of ν_t is a function of the cost overrun and the second part of ν_t is the unobserved heterogeneity ϵ_t. We exploit this decomposition and assume that these two error terms are conditionally independent given (Z,θ). This is equivalent to assuming that the observed cost C and ϵ_t are conditionally independent given (θ,Z) or equivalently (P,Z). If we adopt the second interpretation in subsection 3.1 for introducing ϵ_t, then ϵ_t can be interpreted as the *residual transfer* upon the regulator's discretion. Because such a residual transfer is unrelated to cost efficiency, it is natural to assume that C and ϵ_t are independent conditional upon (θ,Z). Moreover, this assumption would lead to a simple way to identify $\mu(\cdot)$ as shown in the next proposition. For this reason and the fact that it does not require any instrument, we choose this identifying assumption.

Assumption C2: *C and ϵ_t are independent conditional upon (θ, Z).*[31]

In particular, because $P = p^*(\theta, Z)$ and $e = e^*(\theta, Z)$, it follows from (20)-(21) that assumption C2 is satisfied if (ϵ_d, ϵ_c) is independent of ϵ_t given (θ, Z). The next proposition establishes the nonparametric identification of the cost of public funds $\lambda(\cdot)$ from observations (Y, C, P, T, Z).

Proposition 7: *In the basic model with assumptions A1, B1–B3, and C1-C2, the cost-of-public-funds function $\lambda(\cdot)$ is uniquely determined by $\lambda(z) = \mu(z)/[1 - \mu(z)]$, where*

$$\mu(z) = \frac{1}{\mathrm{E}[Y|P=p, Z=z]} \left\{ \frac{\partial \mathrm{E}[T|P=p, Z=z]}{\partial p} \frac{\mathrm{Var}[C|P=p, Z=z]}{\mathrm{Cov}[C, T|P=p, Z=z]} + \frac{\partial \mathrm{E}[C|P=p, Z=z]}{\partial p} - p \frac{\partial \mathrm{E}[Y|P=p, Z=z]}{\partial p} \right\} \tag{46}$$

with $\mathrm{Cov}[C, T|P=p, Z=z] < 0$*, for every* $p \in [\underline{p}(z), \overline{p}(z)]$.

Because p can be chosen arbitrarily, (46) shows that $\mu(\cdot)$ and hence $\lambda(\cdot)$ are overidentified. Thus, weaker assumptions than assumption C2 can be exploited to achieve identification of the cost of public funds. For instance, we could assume that assumption C2 holds for the most efficient firm only. In this case, (46) holds only at $P = \underline{p}(z)$.

5 The General Model

So far we have studied the identification of the basic model with $H(\cdot)$ being the identity function. In this section, we consider the general model. We distinguish two cases depending on whether the function $H(\cdot)$ is known.

5.1. Identification When $H(\cdot)$ Is Known

Hereafter, assumptions A1 and B1 hold. The function $H(\cdot,\cdot)$ takes a known form $H_o(\cdot,\cdot)$ with $H_o'(\cdot,\cdot) > 0$ and $H_o''(\cdot,\cdot) \geq 0$ following assumption A2-(iii), where the prime denotes the derivative with respect to the first argument as before. The function $H_o(\cdot,\cdot)$

[31] As a matter of fact, we only need that cost and residual transfer are uncorrelated given (θ, Z), i.e. $\mathrm{E}[C\epsilon_t|\theta, Z] = 0$. In line with assumption B1, which can be weakened as indicated in footnote 21, we use again a conditional independence requirement that is stronger.

is frequently chosen to be the exponential function in view of the popular Cobb-Douglas specification, in which case efficiency is neutral with respect to all inputs.[32]

The identification of the model with $H_o(\cdot,\cdot)$ known follows a similar path as for the basic model.[33] In view of the proofs of Lemma 4 and Proposition 4, some normalizations are needed for the cost efficiency level of the most efficient firm to recover $\overline{c}_o(p,z)$ and the effort level of the least efficient firm to recover $e^*(p,z)$. Thus we make a similar location-scale normalization.

Assumption D1: *For every value $z \in \mathcal{Z}$*

$$H_o[\underline{\theta}(z) - e^*(\underline{\theta}(z), z)] = 1 \quad \text{and} \quad e^*[\overline{\theta}(z), z] = 0. \tag{47}$$

Since the term $H_o(\cdot,\cdot)$ defines the firm's cost efficiency level, it is natural to set it at one for the most efficient firm and to set the effort for the least efficient firm at zero as discussed in Section 3.2.

We can now discuss the identification of the model. It is straightforward to see that $\overline{y}(p,z) = \mathrm{E}[Y|P=p, Z=z]$. Regarding $\overline{c}_o(p,z)$, we use $\mathrm{E}[C|P=p,Z=z] = H_o(\theta - e, z)\overline{c}_o(p,z)$ and the FOC $p\overline{y}'(p,z) + \mu(z)\overline{y}(p,z) = H_o(\theta-e,z)\overline{c}_o'(p,z)$. The ratio of these two equations provides a differential equation whose solution is given in (29) using the first normalization in (47). On the other hand, instead of (30), the asymmetric information term becomes $\theta - e^*(\theta,z) = \Delta(p,z) \equiv H_o^{-1}\Big[\mathrm{E}[C|P=p,Z=z]/\overline{c}_o(p,z)\Big]$, where the function $\Delta(\cdot,\cdot)$ satisfies $\Delta(\cdot,\cdot) \geq H_o^{-1}(1,\cdot)$ and $\partial\Delta(\cdot,\cdot)/\partial p > 0$. The function $\Gamma(\cdot,\cdot)$ is defined as in (31), while the function $R(\cdot,\cdot)$ is defined as in (32) except that the denominator is replaced by $\Gamma(p,z) - H_o'(\theta-e,z)\overline{c}_o(p,z) + \mu(z)[G_{P|Z}(p|z)/g_{P|Z}(p|z)] \times \partial\Gamma(p,z)/\partial p$. Note that $R(\cdot,\cdot)$ is a function of observables as the term $\theta - e$ can be replaced by $H_o^{-1}\Big[\mathrm{E}[C|P=p,Z=z]/\overline{c}_o(p,z)\Big]$. Hence, the effort disutility is still given by (33) using the second normalization in (47), while the optimal effort and type functions are given by

[32] Gagnepain and Ivaldi (2002), Perrigne (2002) and Brocas, Chan and Perrigne (2006) consider another economic interpretation of the term $\theta - e$. In particular, these papers consider that $\theta - e$ or θ affects labor efficiency, where the effective quantity of labor is (say) $L^* = L/(\theta - e)$ or L/θ. In their parametric specifications, this interpretation constrains the value of the coefficient for $\theta - e$.

[33] To shorten this section, we have avoided to provide formal lemmas and propositions with their proofs. Such proofs are available upon request to the authors.

(34) and (35). Thus the function $F(\cdot|\cdot)$ is recovered as the distribution of θ as previously. As expected, Lemma 5 and Proposition 5 still hold under assumption B3 establishing the identification of $y(\cdot,\cdot,\cdot)$, $c_o(\cdot,\cdot,\cdot)$ and $G_{\epsilon_d,\epsilon_c,\epsilon_t|Z}(\cdot,\cdot,\cdot|\cdot)$. Regarding Lemma 8, the slope of the transfer takes a slightly more involved form because $H_o'(\cdot,\cdot)$ is no longer equal to one. In particular, the slope should be divided by $\partial H_o^{-1}\big[\mathrm{E}[C|P=p,Z=z]/\overline{c}_o(p,z)\big]/\partial p$. Under assumption B4 the cost of public funds is identified, though $\mu(\cdot)$ cannot be obtained explicitly as in (46). Consequently, an algorithm is needed to determine $\mu(\cdot)$.

The following assumption provides some conditions that the distribution of (Y,C,P,T) given Z must satisfy.

Assumption D2: *The joint distribution of (Y,C,P,T) given Z and the cost of public funds $\lambda(\cdot)$ satisfy assumption C1 with the exception of the second inequality in item (iii), which is replaced by $\Gamma(p,z) < H_o'(\Delta(p,z),z)\overline{c}_o(p,z)$.*

Following its proof, Lemma 6 extends to the case when $H_o(\cdot,\cdot)$ is known but not necessarily the identity, namely if a structure $S=\{[y,c_o,H_o,\psi,F,G,\lambda]\}$ satisfies assumptions A1–A3, B1, B3, and D1 then its induced distribution for (Y,C,P,T) given Z that satisfies assumption D2.[34]

5.2. Nonidentification of the General Model

We now consider the general model satisfying assumption A1, where $H(\cdot,\cdot)$ is unknown. As expected, this model also requires a location-scale normalization as in assumption D1 with $H(\cdot,\cdot)$ replacing $H_o(\cdot,\cdot)$. To see this, we can follow a similar argument as in the proof of Lemma 3 by considering a linear transformation of the types and structures satisfying $\tilde{y}(\cdot,\cdot,\cdot)=y(\cdot,\cdot,\cdot)$, $\tilde{c}_o(\cdot,\cdot,\cdot)=c_o(\cdot,\cdot,\cdot)$, $\tilde{H}(\cdot,\cdot)=H(\cdot/\beta,\cdot)$, $\tilde{\psi}(\cdot,\cdot)=\psi((\cdot-\alpha)/\beta,\cdot)$, $\tilde{F}(\cdot)=F((\cdot-\alpha)/\beta)$, $\tilde{G}(\cdot,\cdot,\cdot|\cdot)=G(\cdot,\cdot,\cdot|\cdot)$, and $\tilde{\lambda}(\cdot)=\lambda(\cdot)$.

More importantly, the next proposition shows that one can always find a structure with the identity function for $H(\cdot,\cdot)$ observationally equivalent to the structure $[y,c_o,H,\psi,F,G,\lambda]$. Hence the general model is not identified as one can always rationalize the observations by another structure of the basic model.

Proposition 8: *If $S\in\mathcal{S}'\equiv\{[y,c_o,H,\psi,F,G,\lambda]:$ assumptions A1 – A3 and B1, B3, D1 hold$\}$, then there exists an observationally equivalent structure $\tilde{S}\in\mathcal{S}$ with $H(\cdot,\cdot)$ equal to*

[34] The problem of characterizing all the restrictions imposed by the model with $H_o(\cdot,\cdot)$ known on observables is investigated in Perrigne and Vuong (2007a).

the identity function leading to observations (Y, C, P, T) *given* Z *that satisfy assumption C1.*

Proposition 8 indicates that the general model is not identified. This is not surprising in itself. The results in Sections 3 and 4 suggest that identifying $H(\cdot, \cdot)$ is difficult. The most interesting part of the above result is that one can always explain equivalently the observations with the basic model and a fortiori a model where $H_o(\cdot, \cdot)$ is chosen arbitrarily. The proof of Proposition 8 suggests that there is some "compensation" between the cost efficiency term and the other functions in the model leading to the nonidentification of the general model. Given that a model with $H_o(\cdot, \cdot)$ known is identified, the analyst could entertain several $H_o(\cdot, \cdot)$ functions and empirically assesses the one providing the most economically sensible results.

6 Conclusion

This paper establishes the nonparametric identification of the incentive regulation model under incomplete information, namely adverse selection and moral hazard. We consider a model with stochastic demand and cost functions and a private good. We consider a multiplicatively separable cost function in the base cost and the firm's cost inefficiency. We first show that a location-scale normalization is needed for the cost efficiency of the most efficient firm and the effort level of the least efficient firm. Exploiting the conditional independence of the error terms and the firm's type and the bijective mapping between the observed price and the firm's unknown type, we show that at a given cost of public funds we can recover the structure of the model, namely the demand and base cost functions, the effort disutility function, the distribution of the firms' type and the joint conditional distribution of the stochastic shocks. To identify the cost of public funds, we assume some conditional independence of the unobserved heterogeneity term affecting the transfer with the observed cost. To this end, we derive the restrictions that must be satisfied by the cost of public funds and the observables so as to rationalize the observations by an incentive regulation model. Lastly, we extend our results to a more general model, in which the cost efficiency function is known. When such a function is unknown, the model is nonidentified. Moreover, we show that the latter model is observationally equivalent to a model in which the cost efficiency function is the identity.

Our paper represents a stepping stone in the structural analysis of data subject to incomplete information such as in contracts, insurance and nonlinear pricing. Our identification results indicate that one does not have to rely on parametric functional forms to estimate such models as our analysis is sufficiently general to be extended to other models of incomplete information. As a matter of fact, the incentive regulation model we consider includes some functions such as the effort disutility and the cost of public funds that do not appear in a standard model of incomplete information under adverse selection only such as in nonlinear pricing models. See Huang, Perrigne and Vuong (2007) for the nonparametric identification and estimation of nonlinear pricing models.

Clearly, the problem of estimating and testing such models needs to be addressed. It includes two important questions. First, we need to derive the restrictions imposed by the model on observables to test its validity. Our results (assumption C1, Lemmas 6 and 7) provide a first step toward this goal. Second, incomplete information is generally assumed. It would be interesting to assess the adequacy of such a statement. The restrictions imposed by a complete information model would allow us to test which model (incomplete or complete information) is the most accurate to explain the data. The problem of testing adverse selection has known a vivid interest recently and some tests have been developed within the reduced form approach. See Chiappori and Salanié (2000). A test based on the theoretical restrictions imposed by the model on observables would provide a complete answer to this question within the structural approach. Lastly, it remains to develop suitable nonparametric estimators and to study their asymptotic properties. Our results show how to express the structural elements of the model from the reduced form probability distribution of the observables through various conditional expectations. Thus a multistep estimation procedure could be entertained. A difficulty relies in that many conditional expectations need to be estimated at some boundaries.

Appendix A

This appendix gives the proofs of the propositions and lemmas stated in Section 2.

Proof of Proposition 1: From (8) the Hamiltonian of the optimization problem (P') is

$$\begin{aligned}\mathcal{H} &= \Big\{ \int_p^\infty \overline{y}(v)dv + (1+\lambda)\Big(p\overline{y}(p) - \psi(e) - \mathrm{E}_\epsilon\Big[c(y(p,\epsilon_d),\theta - e,\epsilon_c)\Big]\Big) \\ &\quad -\lambda U(\theta)\Big\} f(\theta) + \gamma(\theta)(-\psi'(e)),\end{aligned}$$

where $p = p(\theta)$ and $e = e(\theta)$ are the control functions, $U(\theta)$ is the state variable, and $\gamma(\theta)$ is the co-state variable. Hence, applying the Pontryagin principle, the FOC are:

$$\begin{aligned}\mathcal{H}_p &= \Big\{\lambda\overline{y}(p) + (1+\lambda)p\overline{y}'(p) - (1+\lambda)\mathrm{E}_\epsilon\Big[c_1(y(p,\epsilon_d),\theta-e,\epsilon_c)y_1(p,\epsilon_d)\Big]\Big\}f(\theta) = 0 \\ \mathcal{H}_e &= \Big\{-(1+\lambda)\psi'(e) + (1+\lambda)\mathrm{E}_\epsilon\Big[c_2(y(p,\epsilon_d),\theta-e,\epsilon_c)\Big]\Big\}f(\theta) - \gamma(\theta)\psi''(e) = 0 \\ -\mathcal{H}_U &= \lambda f(\theta) = \gamma'(\theta).\end{aligned}$$

The last equation gives $\gamma(\theta) = \lambda F(\theta)$ using the transversality condition $\gamma(\underline{\theta}) = 0$. Thus, rearranging $\mathcal{H}_p$ and $\mathcal{H}_e$, the solutions $p = p^*(\theta)$ and $e = e^*(\theta)$ are given by (12) and (13).□

Proof of Proposition 2: Given the price schedule $p^*(\cdot)$ and the transfer function $t^*(\cdot,\cdot)$, we show that the firm will announce its true type θ and exerts the optimal effort $e^*(\theta)$ by verifying the FOC of the firm's problem (F). Under assumption A1, this problem becomes

$$\begin{aligned}(F^*) \qquad \max_{\tilde{\theta},e}\ & \mathrm{E}\Big[t^*\Big(\tilde{\theta}, c(y(p^*(\tilde{\theta}),\epsilon_d),\theta-e,\epsilon_c)\Big) \mid \theta\Big] - \psi(e) \\ &= \mathrm{E}_\epsilon\Big[t^*\Big(\tilde{\theta}, c(y(p^*(\tilde{\theta}),\epsilon_d),\theta-e,\epsilon_c)\Big)\Big] - \psi(e) \\ &= A(\tilde{\theta}) + \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta}-e^*(\tilde{\theta})]}\Big\{H[\tilde{\theta}-e^*(\tilde{\theta})] - H(\theta-e)\Big\} - \psi(e),\end{aligned}$$

where the first equality follows from the independence between θ and (ϵ_d,ϵ_c), while the second equality follows from (15) and (16). Thus, using (17) the FOC with respect to $\tilde{\theta}$ and e are respectively

$$\begin{aligned}0 &= \psi'[e^*(\tilde{\theta})]e^{*\prime}(\tilde{\theta}) - \psi'[e^*(\tilde{\theta})] + \left\{\frac{d}{d\tilde{\theta}}\left(\frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta}-e^*(\tilde{\theta})]}\right)\right\}\Big\{H[\tilde{\theta}-e^*(\tilde{\theta})] - H(\theta-e)\Big\} \\ &\quad + \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta}-e^*(\tilde{\theta})]}H'[\tilde{\theta}-e^*(\tilde{\theta})]\Big[1-e^{*\prime}(\tilde{\theta})\Big]\end{aligned}$$

$$= \left\{ \frac{d}{d\tilde{\theta}} \left(\frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta} - e^*(\tilde{\theta})]} \right) \right\} \left\{ H[\tilde{\theta} - e^*(\tilde{\theta})] - H(\theta - e) \right\}$$

$$0 = \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta} - e^*(\tilde{\theta})]} H'(\theta - e) - \psi'(e).$$

It is easy to see that these FOC are verified if $\tilde{\theta} = \theta$ and $e = e^*(\theta)$.

It remains to show that $[p^*(\cdot), t^*(\cdot, \cdot), e^*(\cdot), U^*(\cdot)]$ solves the FOC of problem (P). In view of the discussion surrounding problem (P'), it suffices to show that the transfer function $t^*(\cdot, \cdot)$ satisfies (6) and (7) where $[p^*(\cdot), e^*(\cdot), U^*(\cdot)]$ solves the FOC of problem (P'). The preceding shows that the transfer function $t^*(\cdot, \cdot)$ satisfies (7). It remains to show that $t^*(\cdot, \cdot)$ also satisfies (6). Using (16), the right-hand side of (6) is

$$\begin{aligned} A(\theta) + \frac{\psi'[e^*(\theta)]}{H'[\theta - e^*(\theta)]} \left\{ H[\theta - e^*(\theta)] - H[\theta - e^*(\theta)] \right\} - \psi[e^*(\theta)] &= A(\theta) - \psi[e^*(\theta)] \\ &= U^*(\theta), \end{aligned}$$

by (14) and (17), as desired.□

Proof of Lemma 1: From the problem (F), the second partial derivative of the firm's objective function with respect to e is

$$\int U_{33}(\tilde{\theta}, \theta, e, \epsilon_d, \epsilon_c) dG(\epsilon_d, \epsilon_c) = \int [t_{22}(\cdot) c_2^2(\cdot) + t_2(\cdot) c_{22}(\cdot)] dG(\epsilon_d, \epsilon_c) - \psi''(e),$$

where we have omitted the arguments of the functions to simplify the notation. But $c_{22}(\cdot) = H''(\theta - e) c_o(y(p(\tilde{\theta}), \epsilon_d), \epsilon_c) \geq 0$ by assumptions A1 and A2-(iii). When the transfer function $t(\cdot, \cdot)$ is weakly decreasing and concave in realized cost c so that $t_2(\cdot) \leq 0$ and $t_{22}(\cdot) \leq 0$, it follows from $\psi''(\cdot) > 0$ that the firm's objective function is *strictly* concave in e for any $(\tilde{\theta}, \theta)$. Hence, the effort $e(\tilde{\theta}, \theta)$, which solves the FOC (3), is uniquely defined and corresponds to a global maximum of the problem (FE).

Next, we show that $0 \leq e_2(\theta, \theta) < 1$. This can be seen by differentiating the FOC (3) defining $e(\tilde{\theta}, \theta)$ with respect to θ. This gives

$$0 = [1 - e_2(\tilde{\theta}, \theta)] \mathrm{E}_\epsilon \Big[t_{22}(\cdot) c_2^2(\cdot) + t_2(\cdot) c_{22}(\cdot) \Big] + \psi''[e(\tilde{\theta}, \theta)] e_2(\tilde{\theta}, \theta).$$

Rearranging and evaluating at $\tilde{\theta} = \theta$ give

$$e_2(\theta, \theta) \Big\{ \mathrm{E}_\epsilon \Big[t_{22}(\cdot) c_2^2(\cdot) + t_2(\cdot) c_{22}(\cdot) \Big] - \psi''[e(\theta)] \Big\} = \mathrm{E}_\epsilon \Big[t_{22}(\cdot) c_2^2(\cdot) + t_2(\cdot) c_{22}(\cdot) \Big].$$

Under assumptions A1 and A2-(iii), we have $c_{22}(\cdot) \geq 0$ as noted above. Thus the expectation term is nonpositive whenever the transfer function $t(\cdot, \cdot)$ is weakly decreasing and concave in realized cost. Because $\psi''(\cdot) > 0$ by assumption A2-(iv), it follows that $0 \leq e_2(\theta, \theta) < 1$.□

Proof of Lemma 2: As noted before assumption A3, the local SOC (19) is satisfied as soon as $e^{*\prime}(\cdot) \leq 0$. We show that $e^{*\prime}(\cdot) < 0$. By definition $[p^*(\cdot), e^*(\cdot)]$ satisfies the FOC (12)-(13), which can be written as

$$p^*(\theta)\overline{y}'[p^*(\theta)] = H[\theta - e^*(\theta)]\overline{c}_o'[p^*(\theta)] - \mu\overline{y}[p^*(\theta)] \quad \text{(A.1)}$$

$$\psi'[e^*(\theta)] = H'[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)] - \mu\frac{F(\theta)}{f(\theta)}\psi''[e^*(\theta)], \quad \text{(A.2)}$$

where we have used assumption A1, the definition of $\overline{c}_o(\cdot)$, and the expression for $\overline{c}_o'(\cdot)$ found earlier. Differentiating (12)-(13) with respect to θ and rearranging equations give

$$Ae^{*\prime}(\theta) + Bp^{*\prime}(\theta) = A \quad \text{(A.3)}$$

$$Ce^{*\prime}(\theta) - Ap^{*\prime}(\theta) = D, \quad \text{(A.4)}$$

where

$$\begin{aligned}
A &= H'[\theta - e^*(\theta)]\overline{c}_o'[p^*(\theta)] \\
B &= (1+\mu)\overline{y}'[p^*(\theta)] + p^*(\theta)\overline{y}''[p^*(\theta)] - H[\theta - e^*(\theta)]\overline{c}_o''[p^*(\theta)] \\
&= (1-\mu)\overline{V}''[p^*(\theta)] - H[\theta - e^*(\theta)]\overline{c}_o''[p^*(\theta)] \\
C &= \psi''[e^*(\theta)] + \mu\frac{F(\theta)}{f(\theta)}\psi'''[e^*(\theta)] + H''[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)] \\
D &= H''[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)] - \mu\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right)\psi''[e^*(\theta)],
\end{aligned}$$

with $\mu = \lambda/(1+\lambda)$. Under assumptions A1–A2, note that $A < 0$, $B < 0$ and $C > 0$. Solving for $e^{*\prime}(\theta)$ gives

$$e^{*\prime}(\theta)\left(C + \frac{A^2}{B}\right) = D + \frac{A^2}{B}.$$

Thus, $e^{*\prime}(\cdot) < 0$ if $-C < A^2/B < -D$, i.e. if

$$\begin{aligned}
&-\left(\psi''[e^*(\theta)] + \mu\frac{F(\theta)}{f(\theta)}\psi'''[e^*(\theta)]\right) \\
&< H''[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)] + \frac{\left\{H'[\theta - e^*(\theta)]\overline{c}_o'[p^*(\theta)]\right\}^2}{(1-\mu)\overline{V}''[p^*(\theta)] - H[\theta - e^*(\theta)]\overline{c}_o''[p^*(\theta)]} \\
&< \mu\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right)\psi''[e^*(\theta)]. \quad \text{(A.5)}
\end{aligned}$$

Because $-B \geq -(1-\mu)\overline{V}''[p^*(\theta)] > 0$, assumption A3-(i) ensures that

$$-\psi''[e^*(\theta)] < \frac{\left\{H'[\theta - e^*(\theta)]\overline{c}_o'[p^*(\theta)]\right\}^2}{(1-\mu)\overline{V}''[p^*(\theta)] - H[\theta - e^*(\theta)]\overline{c}_o''[p^*(\theta)]},$$

which implies the first inequality in (A.5) by assumption A2. Assumption A3-(ii) is equivalent to

$$H''[\theta - e^*(\theta)]\overline{c}_o[p^*(\theta)] \leq \mu \frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \psi''[e^*(\theta)],$$

which implies the second inequality in (A.5) because $B < 0$ and $H'[\theta - e^*(\theta)]\overline{c}^{o\prime}[p^*(\theta)] \neq 0$ under assumption A2.

Lastly, because $e^{*\prime}(\theta) + p^{*\prime}(\theta)B/A = 1$ by (A.3) with $A < 0$ and $B < 0$, it follows from $e^{*\prime}(\cdot) < 0$ that $p^{*\prime}(\cdot) > 0$, as desired.□

Proof of Proposition 3: Recalling that $e(\tilde{\theta}, \theta)$ is the optimal level of effort for a firm with type θ, the firm's expected utility (4) from announcing $\tilde{\theta}$ is

$$U(\tilde{\theta}, \theta) = A(\tilde{\theta}) + \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta} - e^*(\tilde{\theta})]}\left\{H[\tilde{\theta} - e^*(\tilde{\theta})] - H[\theta - e(\tilde{\theta}, \theta)]\right\} - \psi[e(\tilde{\theta}, \theta)]$$

(see the optimization problem (F^*) in the proof of Proposition 2). To show that $\tilde{\theta} = \theta$ provides a global maximum, we first show that $U_{12}(\tilde{\theta}, \theta) > 0$ for any $(\tilde{\theta}, \theta)$. From (18), this is equivalent to showing $e_1(\tilde{\theta}, \theta) < 0$, where $e(\tilde{\theta}, \theta)$ solves the FOC (3), which can be written under assumption A1 as

$$0 = \frac{\psi'[e^*(\tilde{\theta})]}{H'[\tilde{\theta} - e^*(\tilde{\theta})]} H'[\theta - e(\tilde{\theta}, \theta)] - \psi'[e(\tilde{\theta}, \theta)],$$

from the FOC of problem (F^*). Differentiating this FOC with respect to $\tilde{\theta}$ gives

$$e_1(\tilde{\theta}, \theta)\left\{\psi''(\cdot) + \frac{\psi'(\cdot)}{H'(\cdot)}H''(\cdot)\right\} = H'(\cdot)\left\{\frac{\psi''(\cdot)e^{*\prime}(\cdot)}{H'(\cdot)} - \frac{\psi'(\cdot)H''(\cdot)[1 - e^{*\prime}(\cdot)]}{H'^2(\cdot)}\right\}.$$

Because $e^{*\prime}(\cdot) < 0$ by Lemma 2, it is easy to verify that the right-hand side is strictly negative while the term in braces is strictly positive under assumption A2. Hence $e_1(\tilde{\theta}, \theta) < 0$ implying $U_{12}(\cdot, \cdot) > 0$, as desired. Second, we apply the argument in Appendix A1.4 in Laffont and Tirole (1993) with $\phi(\beta, \hat{\beta})$ equal to $U(\tilde{\theta}, \theta)$. This establishes that $\tilde{\theta} = \theta$ provides the global maximum of $U(\tilde{\theta}, \theta)$.

To prove the second part, let $\overline{t}(\theta) \equiv \mathrm{E}_\epsilon\left[t^*\left(\theta, c(y(p^*(\theta), \epsilon_d), \theta - e^*(\theta), \epsilon_c)\right)\right]$ so that $\overline{t} = \overline{t}(\theta)$. Let $H_\dagger(\theta) \equiv H[\theta - e^*(\theta)] = \mathrm{E}_\epsilon\left[c(y(p^*(\theta), \epsilon_d), \theta - e^*(\theta), \epsilon_c)\right]/\overline{c}_o[p^*(\theta)]$ so that the firm's cost inefficiency level H^* satisfies $H^* = H_\dagger(\theta)$. Note that $H_\dagger(\cdot)$ is strictly increasing in θ because $dH_\dagger/d\theta = [1 - e^{*\prime}(\theta)]H'[\theta - e^*(\theta)] > 0$ as $H'(\cdot) > 0$ and $e^{*\prime}(\cdot) < 0$. Thus $\theta = H_\dagger^{-1}(H^*)$. We want to show that $\overline{t}_\dagger(H^*) \equiv \overline{t}[H_\dagger^{-1}(H^*)]$ is strictly decreasing and convex in H^*. From (16) and assumption A1, we have $\overline{t}(\theta) = A(\theta)$. Hence, using (17)

$$\frac{d\overline{t}_\dagger}{dH^*} = \frac{A'(\theta)}{H'_\dagger(\theta)} = -\frac{\psi'[e^*(\theta)]}{H'[\theta - e^*(\theta)]},$$

which is strictly negative. Thus, the expected transfer is strictly decreasing in H^*, as desired. Moreover,

$$\frac{d^2\overline{t}_\dagger}{dH^{*2}} = -\frac{\psi''[e^*(\theta)]e^{*\prime}(\theta)}{H'^2[\theta - e^*(\theta)][1 - e^{*\prime}(\theta)]} + \frac{\psi'[e^*(\theta)]}{H'^2[\theta - e^*(\theta)]}H''[\theta - e^*(\theta)][1 - e^{*\prime}(\theta)].$$

It is easy to see that $d^2\overline{t}_\dagger/dH^{*2} > 0$ under assumption A2 because $e^{*\prime}(\theta) < 0$ by Lemma 2. Thus $\overline{t}_\dagger(\cdot)$ is strictly convex in H^*, as desired.□

Appendix B

This appendix gives the proofs of the propositions and lemmas stated in Sections 3, 4 and 5.

Proof of Lemma 3: Let $\tilde{Y}, \tilde{C}, \tilde{P}, \tilde{T}$ denote the endogenous variables under the structure $\tilde{\mathcal{S}}$. Let $\tilde{\theta} \equiv \alpha + \beta\theta$ so that $\tilde{\theta}$ is distributed as $\tilde{F}(\cdot|\cdot) = F[(\cdot - \alpha)/\beta|\cdot]$ conditional upon Z. Let $(\tilde{\epsilon_d}, \tilde{\epsilon_c}, \tilde{\epsilon_t}) \equiv (\epsilon_d, \epsilon_c, \epsilon_t)$ so that $(\tilde{\epsilon_d}, \tilde{\epsilon_c}, \tilde{\epsilon_t})$ is jointly distributed as $\tilde{G}(\cdot,\cdot,\cdot|\cdot) = G(\cdot,\cdot,\cdot|\cdot)$ conditional upon Z. We show that $(\tilde{Y}, \tilde{C}, \tilde{P}, \tilde{T}) = (Y, C, P, T)$, which implies the desired result.

Using $\tilde{y}(\cdot,\cdot,\cdot) = y(\cdot,\cdot,\cdot)$, $\tilde{c}_o(\cdot,\cdot,\cdot) = c_o(\cdot,\cdot,\cdot)/\beta$, (25) and (26), we note that

$$\overline{\tilde{y}}(\cdot,\cdot) = \int \tilde{y}(\cdot,\cdot,\tilde{\epsilon}_d)d\tilde{G}(\tilde{\epsilon}_d|\cdot) = \int y(\cdot,\cdot,\epsilon_d)dG(\epsilon_d|\cdot) = \overline{y}(\cdot,\cdot), \tag{B.1}$$

$$\overline{\tilde{c}}_o(\cdot,\cdot) = \int \tilde{c}_o[\tilde{y}(\cdot,\cdot,\tilde{\epsilon}_d),\cdot,\tilde{\epsilon}_c]d\tilde{G}(\tilde{\epsilon}_d,\tilde{\epsilon}_c|\cdot) = \frac{1}{\beta}\int c_o[y(\cdot,\cdot,\epsilon_d),\cdot,\epsilon_c]dG(\epsilon_d,\epsilon_c|\cdot) = \frac{1}{\beta}\overline{c}_o(\cdot,\cdot). \tag{B.2}$$

Now, we consider the FOC (22)–(23) for $(\tilde{P}, \tilde{e})$ and use (B.1)–(B.2), $\tilde{\psi}(\cdot,\cdot) = \psi[(\cdot - \alpha)/\beta, \cdot]$, $\tilde{\theta} = \alpha + \beta\theta$, $\tilde{F}(\cdot|\cdot) = F[(\cdot - \alpha)/\beta|\cdot]$, $\tilde{f}(\cdot|\cdot) = (1/\beta)f[(\cdot - \alpha)/\beta|\cdot]$ and $\tilde{\lambda}(\cdot) = \lambda(\cdot)$ to obtain

$$\begin{aligned}
\tilde{P}\overline{y}'(\tilde{P}, Z) + \mu\overline{y}(\tilde{P}, Z) &= \left(\theta - \frac{\tilde{e} - \alpha}{\beta}\right)\overline{c}'_o(\tilde{P}, Z) \\
\psi'\left(\frac{\tilde{e} - \alpha}{\beta}, Z\right) + \mu\frac{F(\theta|Z)}{f(\theta|Z)}\psi''\left(\frac{\tilde{e} - \alpha}{\beta}, Z\right) &= \overline{c}_o(\tilde{P}, Z),
\end{aligned}$$

since $H(x, z) = x$ and $H'(x, z) = 1$ in the basic model. From the solution $p^*(\theta, z)$ and $e^*(\theta, z)$ of (22)–(23), it follows that $\tilde{P} = p^*(\theta, Z) = P$ and $(\tilde{e} - \alpha)/\beta = e^*(\theta, Z) = e$. In particular, the latter implies that $\tilde{e}^*(\tilde{\theta}, Z) = \alpha + \beta e^*[(\tilde{\theta} - \alpha)/\beta, Z]$. Moreover, because $\tilde{Y} = \tilde{y}(\tilde{P}, Z, \tilde{\epsilon}_d) = y(\tilde{P}, Z, \tilde{\epsilon}_d)$, we obtain $\tilde{Y} = Y$ since $\tilde{P} = P$ and $\tilde{\epsilon}_d = \epsilon_d$.

Next, we turn to cost and transfer. From (21) with $H(x,z)=x$, and using (B.2), $\tilde{\theta}=\alpha+\beta\theta$ and $\tilde{e}=\alpha+\beta e$, we have

$$\tilde{C}=(\tilde{\theta}-\tilde{e})\tilde{c}_o(\tilde{Y},Z,\tilde{\epsilon}_c)=(\theta-e)c_o(Y,Z,\epsilon_c)=C,$$

since $\tilde{Y}=Y$ and $\tilde{\epsilon}_c=\epsilon_c$. Moreover, from (23) and the previous results we obtain

$$\begin{aligned}
\tilde{T} &= \tilde{\psi}(\tilde{e},Z)+\int_{\tilde{\theta}}^{\overline{\tilde{\theta}}(Z)}\tilde{\psi}'[\tilde{e}^*(\tilde{u},Z),Z]d\tilde{u}-\tilde{\psi}'(\tilde{e},Z)\left\{\frac{\tilde{C}}{\overline{\tilde{c}}_o(\tilde{P},Z)}-(\tilde{\theta}-\tilde{e})\right\}+\tilde{\epsilon}_t\\
&= \psi(e,Z)+\int_{\alpha+\beta\theta}^{\alpha+\beta\overline{\theta}(Z)}\frac{1}{\beta}\psi'\left[e^*\left(\frac{\tilde{u}-\alpha}{\beta},Z\right),Z\right]d\tilde{u}-\frac{1}{\beta}\psi'(e,Z)\left\{\frac{\beta\tilde{C}}{\overline{c}_o(\tilde{P},Z)}-\beta(\theta-e)\right\}+\tilde{\epsilon}_t\\
&= \psi(e,Z)+\int_{\theta}^{\overline{\theta}(Z)}\psi'\left[e^*(u,Z),Z\right]du-\psi'(e,Z)\left\{\frac{\tilde{C}}{\overline{c}_o(\tilde{P},Z)}-(\theta-e)\right\}+\tilde{\epsilon}_t,
\end{aligned}$$

where the second equality uses (B.2), $\tilde{\psi}(\cdot,\cdot)=\psi[(\cdot-\alpha)/\beta,\cdot]$, $\tilde{\theta}=\alpha+\beta\theta$, $\tilde{e}=\alpha+\beta e$ and $\tilde{e}^*(\tilde{\theta},Z)=\alpha+\beta e^*[(\tilde{\theta}-\alpha)/\beta,Z]$, while the third equality follows from the change of variable $u=(\tilde{u}-\alpha)/\beta$. Thus, (23) implies that $\tilde{T}=T$ since $\tilde{C}=C$, $\tilde{P}=P$ and $\tilde{\epsilon}_t=\epsilon_t$.

Lastly, because of the linear transformation given every value of Z, it is easy to verify that the structure $\tilde{\mathcal{S}}$ satisfies assumptions A1–A3 and B1 as soon as the structure $\mathcal{S}$ satisfies these assumptions.□

Proof of Lemma 4: Recall that $P=p^*(\theta,Z)$, where $p^*(\cdot,\cdot)$ is the optimal price schedule. Assumption B1 implies that P is independent of ϵ_d given Z. Hence, (20) gives $\mathrm{E}[Y|P=p,Z=z]=\mathrm{E}[y(p,z,\epsilon_d)|P=p,Z=z]=\mathrm{E}[y(p,z,\epsilon_d)|Z=z]=\int y(p,z,\epsilon_d)dG(\epsilon_d|z)=\overline{y}(p,z)$ by (25). This establishes (28).

Regarding (26), we recall that θ can be expressed as a function $\theta^*(P,Z)=p^{*-1}(P,Z)$, which is strictly increasing in P since $P=p^*(\theta,Z)$ is strictly increasing in θ by Lemma 2. Thus, e can be expressed as a function $e^*(P,Z)$, which is strictly decreasing in P because $e=e^*(\theta,Z)$ is strictly decreasing in θ by Lemma 2, while $\theta=\theta^*(P,Z)$ is strictly increasing in P. Now, from (21) with $H(x,z)=x$ and using $\theta-e=\theta^*(P,Z)-e^*(P,Z)$, we obtain

$$\begin{aligned}
\mathrm{E}[C|P=p,Z=z] &= (\theta-e)\mathrm{E}\Big[c_o[y(p,z,\epsilon_d),z,\epsilon_c]|P=p,Z=z\Big]\\
&= (\theta-e)\mathrm{E}\Big[c_o[y(p,z,\epsilon_d),z,\epsilon_c]|Z=z\Big]\\
&= (\theta-e)\int c_o[y(p,z,\epsilon_d),z,\epsilon_c]dG(\epsilon_d,\epsilon_c|z)\\
&= (\theta-e)\overline{c}_o(p,z), \qquad\qquad (\mathrm{B.3})
\end{aligned}$$

where $\theta-e=\theta^*(p,z)-e^*(p,z)$. The second equality follows from assumption B1 and the last equality follows from (26). In particular, (B.3) establishes (30) with $\Delta(\cdot,\cdot)$ satisfying $\Delta(\cdot,\cdot)\geq 1$

and $\partial\Delta(\cdot,\cdot)/\partial p > 0$ because $\theta - e = \theta^*(p,z) - e^*(p,z)$ is strictly increasing in p with strictly positive derivative with respect to p by Lemma 2 implying $\theta^*(p,z) - e^*(p,z) \geq \theta^*[\underline{p}(z),z] - e^*[\underline{p}(z),z] = \underline{\theta}(z) - e^*[\underline{\theta}(z),z] = 1$ by (27). Moreover, writing (B.3) at $p = \underline{p}(z)$, which is the price for the most efficient firm with type $\underline{\theta}(z)$ and exerting the maximal effort $\overline{e}(z) = e^*[\underline{\theta}(z),z]$, we obtain

$$\mathrm{E}[C|P = \underline{p}(z), Z = z] = \overline{c}_o[\underline{p}(z),z], \tag{B.4}$$

because $[\underline{\theta}(z) - \overline{e}(z)] = 1$ by the normalization (27).

Next, we write (22) with $H(x,z) = x$ at $Z = z$ so that $P = p^*(\theta,z) = p$ and $e = e^*(\theta,z)$. Dividing the resulting equation by (B.3) we obtain

$$\frac{p\overline{y}'(p,z) + \mu\overline{y}(p,z)}{\mathrm{E}[C|P = p, Z = z]} = \frac{\overline{c}_o'(p,z)}{\overline{c}_o(p,z)}.$$

Integrating this differential equation from $\underline{p}(z)$ to some arbitrary $p \in [\underline{p}(z), \overline{p}(z)]$, where $\overline{p}(z) \equiv p^*[\overline{\theta}(z),z)]$ is the price for the least efficient type when $Z = z$, we obtain

$$\log\left(\frac{\overline{c}_o(p,z)}{\overline{c}_o(\underline{p}(z),z)}\right) = \int_{\underline{p}(z)}^{p} \frac{\tilde{p}\overline{y}'(\tilde{p},z) + \mu\overline{y}(\tilde{p},z)}{\mathrm{E}[C|P = \tilde{p}, Z = z]} d\tilde{p}.$$

Solving for $\overline{c}_o(p,z)$ and using the boundary condition (B.4) give (29).□

Proof of Proposition 4: Because $\theta - e = \theta^*(p,z) - e^*(p,z)$, differentiating (30) with respect to p gives

$$\frac{\partial\theta^*(p,z)}{\partial p} - \frac{\partial e^*(p,z)}{\partial p} = \frac{\partial\Delta(p,z)}{\partial p} > 0, \tag{B.5}$$

where $\partial\theta^*(\cdot,\cdot)/\partial p > 0$ and $\partial e^*(\cdot,\cdot)/\partial p < 0$ from Lemma 2. In particular, $\theta = \theta^*(P,Z)$ and $e = e^*(P,Z)$ are in bijections with P given Z. Thus, taking conditional expectation of (24) given $(P,Z) = (p,z)$, and using (B.3) together with $\mathrm{E}[\epsilon_t|P = p, Z = z] = \mathrm{E}[\epsilon_t|\theta, Z = z] = \mathrm{E}[\epsilon_t|Z = z] = 0$ by assumption B1, we obtain

$$\mathrm{E}[T|P{=}p, Z{=}z] = \psi(e,z) + \int_{\theta}^{\overline{\theta}(z)} \psi'[e^*(\tilde{\theta},z),z]d\tilde{\theta}, \tag{B.6}$$

where $\theta = \theta^*(p,z)$, $\overline{\theta}(z) = \theta^*[\overline{p}(z),z]$ and $e = e^*(p,z)$. Differentiating (B.6) gives

$$\frac{\partial\mathrm{E}[T|P{=}p, Z{=}z]}{\partial p} = \psi'(e,z)\left(\frac{\partial e^*(p,z)}{\partial p} - \frac{\partial\theta^*(p,z)}{\partial p}\right), \tag{B.7}$$

where we have used $e^*(\theta,z) = e^*(p,z) = e$. Thus, (B.5), (B.7) and (31) give

$$\psi'(e,z) = \Gamma(p,z) > 0, \tag{B.8}$$

because $\psi'(\cdot, z) > 0$ by assumption A2. Differentiating (B.8) again gives

$$\psi''(e,z)\frac{\partial e^*(p,z)}{\partial p} = \frac{\partial \Gamma(p,z)}{\partial p} < 0, \tag{B.9}$$

because $\psi''(\cdot, z) > 0$ by assumption A2 and $\partial e^*(p,z)/\partial p < 0$ by Lemma 2. Using (B.8)–(B.9) into (23) with $H'(x,z) = 1$ at $Z = z$ so that $P = p^*(\theta, z) = p$ and $e = e^*(\theta, z)$, we obtain

$$\Gamma(p,z) + \mu \frac{G_{P|Z}(p,z)}{g_{P|Z}(p,z)} \frac{\partial \Gamma(p,z)}{\partial p} \frac{\partial \theta^*(p,z)/\partial p}{\partial e^*(p,z)/\partial p} = \overline{c}_o(p,z), \tag{B.10}$$

where we have used the property that $F(\theta|z)/f(\theta|z) = [G_{P|Z}(p,z)/g_{P|Z}(p,z)]\partial\theta^*(p,z)/\partial p$ because $\theta = \theta^*(p,z)$ is strictly increasing in p from Lemma 2.

We now solve (B.5) and (B.10) for $\partial e^*(p,z)/\partial p$ and $\partial \theta^*(p,z)/\partial p$ to obtain after some algebra

$$\frac{\partial e^*(p,z)}{\partial p} = -R(p,z) < 0 \tag{B.11}$$

$$\frac{\partial \theta^*(p,z)}{\partial p} = \frac{\partial \Delta(p,z)}{\partial p} - R(p,z) > 0, \tag{B.12}$$

where $R(p,z)$ is as given in (32) with $R(p,z) > 0$ because $\partial e^*(\cdot,z)/\partial p < 0$ by Lemma 2. Similarly, the right-hand side of (B.12) must be strictly positive because $\partial\theta^*(\cdot,z)/\partial p > 0$ by Lemma 2 leading to $\partial\Delta(p,z)/\partial p > R(p,z)$. Now, note that $e^*[\overline{p}(z), z] = e^*[\overline{\theta}(z), z] = 0$ by (27). Moreover, from (30), we have $\theta^*[\overline{p}(z), z] - 0 = \Delta[\overline{p}(z), z]$. Thus, integrating (B.11) and (B.12) from some arbitrary $p \in [\underline{p}(z), \overline{p}(z)]$ to $\overline{p}(z)$, and using the preceding boundary conditions, we obtain (34) and (35). As all the functions on the right-hand side of (35) are identified, it follows that the firm's type θ can be recovered from (p, z), and that the conditional distribution $F(\cdot|z)$ of type is identified as the distribution of $\theta = \theta^*(P, z)$, where P is distributed as $G_{P|Z}(\cdot|z)$.

Lastly, let $\underline{e}(z) \equiv e^*[\overline{p}(z), z] = e^*[\overline{\theta}(z), z] = 0$ by the normalization (27), and let $\overline{e}(z) \equiv e^*[\underline{p}(z), z] = e^*[\underline{\theta}(z), z]$. Integrating (B.8) from 0 to some arbitrary $e \in [\underline{e}(z), \overline{e}(z)]$, where $p = p^*(\cdot, z)$ is the inverse function of $e^*(\cdot, z)$, gives

$$\psi(e,z) = \psi(0,z) + \int_0^e \Gamma[p^*(\tilde{e}, z), z] d\tilde{e},$$

which establishes (33) since (B.6) evaluated at $p = \overline{p}(z)$ gives $\mathrm{E}[T|P = \overline{p}(z), Z = z] = \psi(0, z)$ as $e = e^*[\overline{p}(z), z] = 0$ and $\theta = \theta^*[\overline{p}(z), z] = \overline{\theta}(z)$ when $p = \overline{p}(z)$.□

Proof of Lemma 5: Let $\tilde{\epsilon}_d = y(p_o(Z), Z, \epsilon_d)$ for an arbitrary $p_o(\cdot) \in [\underline{p}(\cdot), \overline{p}(\cdot)]$. Thus, $\epsilon_d = y^{-1}[p_o(Z), Z, \tilde{\epsilon}_d]$ since $y(\cdot,\cdot,\cdot)$ is strictly increasing in ϵ_d by assumption. Moreover, let $y(p, z, y^{-1}[p_o(z), z, \tilde{\epsilon}_d]) = \tilde{y}(p, z, \tilde{\epsilon}_d)$, which is strictly increasing in $\tilde{\epsilon}_d$. Thus, we have $\tilde{y}(p_o(z), z, \tilde{\epsilon}_d)$

$= \tilde{\epsilon}_d$ thereby satisfying the first equality in (39). We need to verify that $\overline{y}(p,z) = \overline{\overline{y}}(p,z)$. In particular,

$$\begin{aligned} \overline{\overline{y}}(p,z) &= \mathrm{E}[\tilde{y}(p,z,\tilde{\epsilon}_d)|p,z] \\ &= \mathrm{E}[y(p,z,y^{-1}[p_0(z),z,\tilde{\epsilon}_d])|p,z] \\ &= \mathrm{E}[y(p,z,\epsilon_d)|p,z] = \overline{y}(p,z). \end{aligned}$$

We can apply the same reasoning for $\tilde{c}_o(y,z,\tilde{\epsilon}_c)$. Let $\tilde{\epsilon}_c = c_o(y_o(Z),Z,\epsilon_c)$ for an arbitrary $y_o(\cdot) \in [\underline{y}(\cdot),\overline{y}(\cdot)]$. Thus, $\epsilon_c = c_o^{-1}[y_o(Z),Z,\tilde{\epsilon}_c]$ since $c_o(\cdot,\cdot,\cdot)$ is strictly increasing in ϵ_c by assumption. Let $c_o(y,z,c_o^{-1}[y_o(z),z,\tilde{\epsilon}_c]) = \tilde{c}_o(y,z,\tilde{\epsilon}_c)$, which is strictly increasing in $\tilde{\epsilon}_c$. Thus, we have $\tilde{c}_o(y_0(z),z,\tilde{\epsilon}_c) = \tilde{\epsilon}_c$ thereby satisfying the second inequality in (39). We need to verify that $\overline{c}_o(p,z) = \overline{\overline{c}}_o(p,z)$. In particular,

$$\begin{aligned} \overline{\overline{c}}_o(y,z) &= \mathrm{E}[\tilde{c}_o(y,z,\tilde{\epsilon}_c)|p,z] \\ &= \mathrm{E}[c_o(\tilde{y}(p,z,\tilde{\epsilon}_d),z,c_o^{-1}[y_o(z),z,\tilde{\epsilon}_c])|p,z] \\ &= \mathrm{E}[c_o(y(p,z,\epsilon_d)z,\epsilon_c)|p,z] = \overline{c}_o(p,z). \end{aligned}$$

It remains to show that these two models are observationally equivalent, which is straightforward. Since $\tilde{\epsilon}_d = y(p_o(Z),Z,\epsilon_d)$, we have $\tilde{y}(P,Z,\tilde{\epsilon}_d) = y(P,Z,y^{-1}[p_o(Z),Z,y(p_o(Z),Z,\epsilon_d)]) = y(P,Z,\epsilon_d)$. Similarly, since $\tilde{\epsilon}_c = c_o(y_o(Z),Z,\epsilon_c)$, we have $\tilde{c}_o(Y,Z,\tilde{\epsilon}_c) = c_o(Y,Z,c_o^{-1}[y_o(Z),Z, c_o(y_o(Z),Z,\epsilon_c)]) = c_o(Y,Z,\epsilon_c)$. Thus, we have $Y = y(P,Z,\epsilon_d) = \tilde{y}(P,Z,\tilde{\epsilon}_d)$ and $C_o = c_o(Y,Z,\epsilon_c)$ $= \tilde{c}_o(Y,Z,\tilde{\epsilon}_c)$. Lastly, P is conditionally independent of $(\tilde{\epsilon}_d,\tilde{\epsilon}_c)$ given Z because P is conditionally independent of (ϵ_d,ϵ_c) given Z combined with $\tilde{\epsilon}_d = y(p_o(Z),Z,\epsilon_d)$ and $\tilde{\epsilon}_c = c_o(y_o(Z),Z,\epsilon_c)$. Because the latter functions are strictly increasing in ϵ_d and ϵ_c, respectively, it follows that $G_{\tilde{\epsilon}_d|Z}(\cdot|\cdot)$ and $G_{\tilde{\epsilon}_c|\tilde{\epsilon}_d},Z(\cdot|\cdot,\cdot)$ are strictly increasing in their first arguments.□

Proof of Proposition 5: Part (i) follows exactly Matzkin (2003) identification argument. Because ϵ_d is independent of θ given Z by assumption B1, then ϵ_d is independent of $P = p^*(\theta,Z)$ given Z. Thus, if $G_{\epsilon_d|P,Z}(\cdot|\cdot,\cdot)$ denotes the conditional distribution of ϵ_d given (P,Z), then for every (p,z) we have $G_{\epsilon_d|Z}(\cdot|z) = G_{\epsilon_d|P,Z}(\cdot|p,z) = G_{Y|P,Z}[y(p,z,\cdot)|p,z]$ because $y(p,z,\cdot)$ is strictly increasing in ϵ_d. In particular, this shows that $G_{Y|P,Z}(\cdot|p,z)$ is strictly increasing in its first argument in view of the second part of assumption B3. Hence, for every (p,z) we have

$$y(p,z,\cdot) = G_{Y|P,Z}^{-1}[G_{\epsilon_d|Z}(\cdot|z)|p,z]. \tag{B.13}$$

Moreover, letting $p = p_o(z)$ we obtain $G_{\epsilon_d|Z}(\cdot|z) = G_{Y|P,Z}[y(p_o(z),z,\cdot)|p_o(z),z] = G_{Y|P,Z}[\cdot|p_o(z), z]$, where the second equality follows from the first normalization in (39). This establishes (41) and hence (40) using (B.13).

To prove (ii) we extend Matzkin's argument as $Y = y(P, Z, \epsilon_d)$ is not independent from ϵ_c given Z in $C_o = c_o(Y, Z, \epsilon_c)$. On the other hand, we exploit the fact that P is independent from ϵ_c given (ϵ_d, Z) because $P = p^*(\theta, Z)$ and θ is independent of ϵ_c given (ϵ_d, Z) by assumption B1. Thus, similarly to above, we obtain $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z) = G_{\epsilon_c|\epsilon_d,P,Z}(\cdot|\epsilon_d, p, z) = G_{C_o|\epsilon_d,P,Z}\{c_o[y(p, z, \epsilon_d), z, \cdot]|\epsilon_d, p, z\} = G_{C_o|Y,P,Z}[c_o(y, z, \cdot)|y, p, z]$ because $c_o(y, z, \cdot)$ is strictly increasing in ϵ_c and $y \equiv y(p, z, \epsilon_d)$. In particular, $G_{C_o|Y,P,Z}(\cdot|y, p, z)$ is strictly increasing in its first argument in view of the second part of assumption B3. Hence, we have

$$c_o(y, z, \cdot) = G^{-1}_{C_o|Y,P,Z}[G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z)|y, p, z], \tag{B.14}$$

for every (y, p, z, ϵ_d) satisfying $y = y(p, z, \epsilon_d)$. We now exploit the second normalization in (39). Given the additional conditional in (ii), let $p = p_\dagger(z, \epsilon_d)$ in (B.14). We obtain

$$\begin{aligned} G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z) &= G_{C_o|Y,P,Z}\Big[c_o\{y[p_\dagger(z), z, \epsilon_d)], z, \cdot\} \mid y[p_\dagger(z), z, \epsilon_d)], p_\dagger(z, \epsilon_d), z\Big] \\ &= G_{C_o|Y,P,Z}\Big[c_o\{y_o(z), z, \cdot\} \mid y_o(z), p_\dagger(z, \epsilon_d), z\Big] \\ &= G_{C_o|Y,P,Z}\left[\cdot \mid y_o(z), p_\dagger, z\right], \end{aligned}$$

where the second equality follows from the additional condition in (ii), while the third equality follows from the second normalization in (39). This establishes (43). Equation (42) follows from (B.14). Lastly, $p_\dagger(\cdot, \cdot)$ is identified as $y_o(\cdot)$ is known and $y(\cdot, \cdot, \cdot)$ is identified by (i).□

Proof of Lemma 6: Hereafter, the results hold for any $p \in [\underline{p}(z), \overline{p}(z)]$ and $z \in \mathcal{Z}$. The first condition in assumption C1-(i) follows from (25) and by using $y(\cdot, \cdot, \cdot) \geq 0$ together with assumption B3-(ii) and the nondegeneracy of the distribution of ϵ_d given Z by assumption B3-(ii). The second condition of assumption C1-(i) follows from $\theta - e > 0$ by assumption A1, and assumption A2-(i). Regarding assumption C1-(ii), following Proposition 4, $\psi'(e, z) = \Gamma(p, z)$. From assumption A2-(iv), $\psi'(e, z) > 0$ leads to $\Gamma(p, z) > 0$. Similarly, $\psi''(e, z) = (\partial\Gamma(p, z)/\partial p)/e'(p, z)$. Since $e'(p, z) < 0$ by Lemma 2 and $\psi''(e, z) > 0$ by assumption A2-(iv), we have $\partial\Gamma(p, z)/\partial p < 0$. Regarding assumption C1-(iii), from Proposition 4, $\theta'(p, z) = \partial\Delta(p, z)/\partial p - R(p, z)$ and by Lemma 2, $p'(\theta, z) > 0$ or equivalently $\theta'(p, z) > 0$. Thus, we have $\partial\Delta(p, z)/\partial p > R(p, z)$, which implies $\Gamma(p, z) < \overline{c}_o(p, z)$ as discussed in the text. Regarding assumption C1-(iv), it is a direct consequence of assumption B1. Given the monotonic relationship between θ and P, the error terms are also conditionally independent of P given Z. The desired result follows from $\epsilon_d = \phi_d(Y, P, Z)$, $\epsilon_c = \phi_c(Y, C, P, Z)$ and $\epsilon_t = \phi_t(Y, C, P, T, Z)$. Lastly, the first part of assumption C1-(v) follows from the assumptions on $F(\cdot|\cdot)$ combined with $P = p(\theta, Z)$ and $p'(\cdot, \cdot) > 0$ by Lemma 2. Regarding the second part of assumption C1-(v), using $Y = y(P, Z, \epsilon_d)$ and $C = (\theta - e)c_o(Y, Z, \epsilon_c)$

combined with assumptions B1 and B3-(ii), we obtain $G_{Y|P,Z}(\cdot|p,z) = G_{\epsilon_d|Z}[y^{-1}(p,z,\cdot)|z]$ and $G_{C|Y,P,Z}(\cdot|y,p,z) = G_{\epsilon_c|\epsilon_d,Z}[c_o^{-1}(y,z,\cdot/(\theta-e))|\epsilon_d,z]$, which establish the desired property. □

Proof of Lemma 7: To prove the first part, we need to show that the conditional distribution of (Y,C,P,T) given Z induced by a structure $S \in \mathcal{S}$ satisfying the conclusions of Lemmas 1-2 and Proposition 3 also satisfy assumption C1. Regarding assumption C1-(i), the first condition follows as in the proof of Lemma 6. The second condition follows by taking the conditional expectation of (21) given (P,Z) and using $\theta - e > 0$ by asssumption A1 and $\overline{c}_o(p,z) > 0$. The latter inequality follows from (26) using $c_o(\cdot,\cdot,\cdot) \geq 0$ together with assumption B3-(ii). Regarding assumption C1-(ii), from the second part of the proof of Proposition 3 with $H(\theta - e,z) = \theta - e$, we have $d\overline{t}_\dagger/dH = -\psi'[e(\theta,z),z] < 0$ implying $\psi'[e(\theta,z),z] > 0$. From (33), $\psi'(e,z) = \Gamma[p(e,z),z]$. Hence, $\Gamma[p(e,z),z] > 0$. Regarding $\Gamma'[p(e,z),z]$, the proof of Lemma 1 shows that the second partial derivative of the firm's objective function is equal to $-\psi''(e)$ because of the linearity in C of the transfer (24) and $H(\theta-e,z) = \theta - e$. Thus, strict concavity implies $\psi''(e) > 0$. As $\psi''(e) = \Gamma'[p(e,z),z]p'(e,z)$ from (33), where $p'(e,z) < 0$ because $p'(\theta,z) > 0$ and $e'(\theta,z) < 0$ by assumption, it follows that $\Gamma'[p(e,z),z] < 0$.

Regarding assumption C1-(iii), we have $\Delta'(p,z) = \theta'(p,z) - e'(p,z) > 0$ because $p'(\theta,z) > 0$ and $e'(p,z) < 0$ by assumption. Moreover, from (35), we have $\theta'(p,z) = \Delta'(p,z) - R(p,z) > 0$ by assumption. Thus $\Delta'(p,z) > R(p,z)$ or equivalently $\Gamma'(p,z) < \overline{c}_o(p,z)$ as discussed in the text. Regarding assumption C1-(iv), from (20), (21), (24) and assumption B3-(ii), the proofs of Lemma 4, Propositions 4 and 5 show that the error terms $(\epsilon_d,\epsilon_c,\epsilon_t)$ can be recovered through identified functions $\phi_d(Y,P,Z)$, $\phi_c(Y,C,P,Z)$ and $\phi_t(Y,P,C,T,Z)$. Since P and θ are in a bijective relationship, the latter are conditionally independent of P given Z in view of assumption B1. Regarding assumption C1-(v), the first part follows from $p'(\theta,z) > 0$ and $f_{\theta|Z}(\cdot|\cdot) > 0$, while the second part follows from assumption B3-(ii), $G_{Y|P,Z}(\cdot|p,z) = G_{\epsilon_d|Z}[y^{-1}(p,z,\cdot)|z]$ and $G_{C|Y,P,Z}(\cdot|y,p,z) = G_{\epsilon_c|\epsilon_d,Z}[c_o^{-1}(y,z,\cdot/(\theta-e))|\epsilon_d,z]$, where the latter uses the bijective mapping bewteen P and θ given z and assumptions B1-B2.

Turning to the second part, let the observations (Y,P,C,T,Z) and a function $\mu(\cdot)$ satisfy assumption C1. We need to define $[y,c_o,\psi,F,G,\mu]$ from the observables and show that these functions satisfy assumptions A1, B1–B3, as well as the conclusions of Lemmas 1 and 2 and Proposition 3. In view of Proposition 5, let

$$\begin{aligned} y(p,z,\epsilon_d) &= G^{-1}_{Y|P,Z}[G_{Y|P,Z}(\epsilon_d|p_o(z),z)|p,z] \\ c_o(y,z,\epsilon_c) &= G^{-1}_{C_o|Y,P,Z}\Big[G_{C_o|Y,P,Z}[\epsilon_c|y_o(z),p_\dagger(z,\epsilon_d),z]|y,p,z\Big], \end{aligned}$$

where $y = y(p,z,\epsilon_d)$ and $C_o = C/\Delta(P,Z)$ with $\Delta(p,z)$ given in (30). Note that assumption

B3-(i) is satisfied by construction. By assumption C1-(v), $y(p, z, \cdot)$ and $c_o(y, z, \cdot)$ are strictly increasing in their last arguments. Thus, we can define

$$\begin{aligned}\epsilon_d &= y^{-1}(P, Z, Y) \\ \epsilon_c &= c_o^{-1}(Y, Z, C_o).\end{aligned}$$

Note that $C = \Delta(P, Z)C_o = \Delta(P, Z)c_o(Y, Z, \epsilon_c)$, where $\Delta(P, Z) = \theta - e$ and $c_o(Y, Z, \epsilon_c) \geq 0$ because $C \geq 0$. Using assumption C1-(i), $\mathrm{E}[C|P, Z] = \Delta(P, Z)\overline{c}_o(P, Z) > 0$ leading to $\Delta(P, Z) > 0$ because $c_o(P, Z) > 0$ by (29). Thus assumption A1 is satisfied. Using assumption C1-(iv), we have $G_{\epsilon_d|Z}(\cdot|z) = G_{\epsilon_d|P,Z}(\cdot|p_o(z), z) = G_{Y|P,Z}(\cdot|p_o(z), z)$. Thus, $G_{\epsilon_d|Z}(\cdot|z)$ is nondegenerated and strictly increasing in its first argument by assumption C1-(v). Similarly, using assumption C1-(iv), we have $G_{\epsilon_c|\epsilon_d,Z}(\cdot|\epsilon_d, z) = G_{\epsilon_c|\epsilon_d,P,Z}(\cdot|\epsilon_d, p_\dagger(z, \epsilon_d), z) = G_{\epsilon_c|Y,P,Z}(\cdot|y_o(z), p_\dagger(z, \epsilon_d), z) = G_{C_o|Y,P,Z}[\cdot|y_o(z), p_\dagger(z, \epsilon_d), z]$, which is nondegenerated and strictly increasing in its first argument by assumption C1-(v) and $C_o = C/\Delta(P, Z)$ with $\Delta(P, Z) > 0$. Hence, assumption B3-(ii) is satisfied.

In view of Proposition 4, let

$$\begin{aligned}\psi(e, z) &= E[T|P = \overline{p}(z), Z = z] + \int_0^e \Gamma[p(\tilde{e}, z)]d\tilde{e} \\ \theta &= \theta(P, Z) \equiv \Delta(P, Z) + \int_P^{\overline{p}(Z)} R(\tilde{p}, Z)d\tilde{p},\end{aligned}$$

where $p(\cdot, z)$ is the inverse of $e(\cdot, z) = \int_{\cdot}^{\overline{p}(z)} R(\tilde{p}, z)d\tilde{p}$, which is strictly decreasing as $R(\cdot, \cdot) > 0$ from assumption C1-(iii). The distribution $F(\theta|Z)$ is then defined as the distribution of $\theta(P, Z)$ given Z. Note that $\theta'(p, z) > 0$ because $\Delta'(p, z) > R(p, z)$ by assumption C1-(iii). Thus $F(\theta|z)$ admits a density, which is strictly positive on $[\underline{\theta}(z), \overline{\theta}(z)]$ as defined in (37) and (38) by assumption C1-(v). Moreover, the conclusions of Lemma 2 (see items (ii)-(iii) preceding Lemma 7) hold from assumptions C1-(iii). For, item (iii) follows from $\theta'(p, z) > 0$ and item (ii) follows from $e'(p, z) = -R(p, z) < 0$ by assumption C1-(iii). In addition, assumption B2 is satisfied as $\underline{\theta}(z) - e(\underline{\theta}(z), z) = \Delta(\underline{p}(z), z) = 1$ by (29) and (30), while $e(\overline{\theta}(z), z) = 0$ because $e(\overline{p}(z), z) = 0$ by construction and $p(\cdot, z)$ strictly increasing leading to $\overline{p}(z) = p(\overline{\theta}, z)$.

Following (24), let

$$\epsilon_t = T - \psi(e, Z) - \int_\theta^{\overline{\theta}(Z)} \psi'(e(\tilde{\theta}, Z), Z)d\tilde{\theta} + \psi'(e, Z)[(C/\overline{c}_o(P, Z)) - \Delta(P, Z)],$$

where $e = e(P, Z)$ and $\psi(\cdot, \cdot)$ are defined as above. Note that $(\epsilon_d, \epsilon_c, \epsilon_t)$ are conditionally independent of P given Z by assumption C1-(iv), and hence of θ given Z. In particular, $\mathrm{E}[\epsilon_t|Z] =$

$\mathrm{E}[\epsilon_t|\overline{p}(Z),Z] = \mathrm{E}[T|\overline{p}(Z),Z] - \psi(e(\overline{\theta}(Z),Z))$ as the third term vanishes when $\theta = \theta[\overline{p}(Z),Z] = \overline{\theta}(Z)$ and the fourth term is equal to zero because $\mathrm{E}[C|\overline{p}(Z),Z] = \Delta(\overline{p}(Z),Z)\overline{c}_o(\overline{p}(Z),Z)$. Hence, assumption B1 is satisfied. Lastly, the conclusions of Lemma 1 and Proposition 3 (see items (i) and (iv)-(v) preceding Lemma 7) hold from assumption C1-(ii). For, with the above transfer in the basic model, the proof of Lemma 1 shows that (i) is equivalent to $\psi''(e,z) > 0$, which is ensured by $\psi''(e,z) = \Gamma'(p,z)e'(p,z)$, $\Gamma'(p,z) < 0$ by assumption C1-(ii) and $e'(p,z) < 0$ as above. In the basic model with the above transfer, the proof of Proposition 3 shows that item (iv) holds when $e'(\theta,z) < 0$, which is established above. Similarly, in the basic model with the above transfer, the proof of Proposition 3 shows that item (v) holds when $\psi'(e,z)) > 0$ and $\psi''(e(\theta,z),z)e'(\theta,z)/[1-e'(\theta,z)] < 0$, which follows from assumption C1-(ii) and the previous results. □

Proof of Proposition 6: Let $S = [y, c_o, \psi, F, G, \lambda]$ be a structure inducing a distribution for (Y,C,P,T) given Z satisfying assumption C1. Define $\tilde{\lambda}(\cdot) = \lambda(\cdot)+\epsilon$, with $\epsilon \neq 0$ sufficiently small so that $\tilde{\lambda}(\cdot)$ and the distribution of (Y,C,P,T) given Z satisfy assumption C1. From Lemma 7, there exists a structure $\tilde{S} = [\tilde{y}, \tilde{c}_o, \tilde{\psi}, \tilde{F}, \tilde{G}, \tilde{\lambda}]$ that satisfies assumptions A1, B1–B3 as well as the conclusions of Lemmas 1 and 2 and Proposition 3 and that rationalizes the observables (Y,C,P,T) given Z. The structure $\tilde{S}$ differs from S because $\tilde{\lambda}(\cdot) \neq \lambda(\cdot)$ though $\tilde{y}(\cdot,\cdot,\cdot) = y(\cdot,\cdot,\cdot)$ and $\tilde{G}_{\epsilon_d|Z}(\cdot|\cdot) = G_{\epsilon_d|Z}(\cdot|\cdot)$. On the other hand, the remaining functions differ.□

Proof of Lemma 8: From (B.8), we have $\psi'(e) = \Gamma[p^*(e,z),z]$, where $e = e^*(\theta,z)$. Hence, from (14) and making the change of variable $\tilde{p} = p^*[e^*(\tilde{\theta},z),z] = p^*(\tilde{\theta},z)$, we obtain

$$U^*(\theta) = \int_p^{\overline{p}(z)} \Gamma(\tilde{p},z)\frac{\partial\theta^*(\tilde{p},z)}{\partial p}d\tilde{p} = \int_p^{\overline{p}(z)} \Gamma(\tilde{p},z)\left[\frac{\partial\Delta(\tilde{p},z)}{\partial p} - R(\tilde{p},z)\right]d\tilde{p},$$

where the second equality follows from (B.12). Then, (45) follows from the participation constraint (10) and

$$\frac{\partial \mathrm{E}[T|P=p,Z=z]}{\partial p} = -\Gamma(p,z)\frac{\partial\Delta(p,z)}{\partial p},$$

which follows from (31).

Similarly, making the change of variable $\tilde{p} = p^*(\tilde{e},z)$ in (33), we obtain

$$\begin{aligned}\psi(e,z) &= \mathrm{E}[T|P=\overline{p}(z),Z=z] + \int_{p^*(0,z)}^{p^*(e,z)} \Gamma(\tilde{p},z)\frac{\partial e^*(\tilde{p},z)}{\partial p}d\tilde{p} \\ &= \mathrm{E}[T|P=\overline{p}(z),Z=z] + \int_p^{\overline{p}(z)} \Gamma(\tilde{p},z)R(\tilde{p},z)d\tilde{p}, \qquad \text{(B.15)}\end{aligned}$$

where the second equality follows from $p^*(0,z) = \overline{p}(z)$ and (B.11). Thus, using (30), (B.6) and (B.8) into (24) we obtain

$$T \;=\; \mathrm{E}[T|P,Z] - \frac{\Gamma(P,Z)}{\overline{c}_o(P,Z)}\Big(C - \mathrm{E}[C|P,Z]\Big) + \epsilon_t. \tag{B.16}$$

We now compute $\Gamma(P,Z)/\overline{c}_o(P,Z)$. From (30), we note that $\Delta(P,Z)\overline{c}_o(P,Z) = \mathrm{E}[C|P,Z]$. Thus, differentiating with respect to p gives

$$\overline{c}_o(P,Z)\frac{\partial\Delta(P,Z)}{\partial p} = \frac{\partial \mathrm{E}[C|P,Z]}{\partial p} - \Delta(P,Z)\frac{\partial \overline{c}_o(P,Z)}{\partial p},$$

where

$$\frac{\partial \overline{c}_o(P,Z)}{\partial p} = \frac{\overline{c}_o(P,Z)}{\mathrm{E}[C|P,Z]}\Big[P\overline{y}'(P,Z) + \mu\overline{y}(P,Z)\Big] = \frac{1}{\Delta(P,Z)}\Big[P\overline{y}'(P,Z) + \mu\overline{y}(P,Z)\Big]$$

from (29)–(30). Hence,

$$\overline{c}_o(P,Z)\frac{\partial\Delta(P,Z)}{\partial p} = \frac{\partial \mathrm{E}[C|P,Z]}{\partial p} - \Big[P\overline{y}'(P,Z) + \mu\overline{y}(P,Z)\Big].$$

It follows from (31) that

$$\frac{\Gamma(P,Z)}{\overline{c}_o(P,Z)} = -\frac{\partial \mathrm{E}[T|P,Z]/\partial p}{\partial \mathrm{E}[C|P,Z]/\partial p - [P\overline{y}'(P,Z) + \mu\overline{y}(P,Z)]}.$$

This establishes (44) in view of (B.16). Lastly, $\mathrm{E}[\epsilon_t|P,Z] = \mathrm{E}[\epsilon_t|Z] = 0$ by Assumption B1. □

Proof of Proposition 7: Because θ and P are in a bijection given Z, we have $\mathrm{E}[C\epsilon_t|P,Z] = \mathrm{E}[C\epsilon_t|\theta,Z] = \mathrm{E}[C|\theta,Z]E[\epsilon_t|\theta,Z] = 0$, where the second equality follows from assumption B4, while the third equality follows from $\mathrm{E}[\epsilon_t|\theta,Z] = 0$ by assumption B1. We then multiply (44) by C and take the expectation of the resulting equation conditional on (P,Z). Using $\mathrm{E}[C\epsilon_t|P,Z] = 0$, it gives

$$\begin{aligned}\mathrm{E}\{C(T - \mathrm{E}[T|P,Z])|P,Z\} \;&=\; -\frac{\partial \mathrm{E}[T|P,Z]/\partial p}{\partial \mathrm{E}[C|P,Z]/\partial p - [P\overline{y}'(P,Z) + \mu\overline{y}(P,Z)]} \\ &\qquad \times \Big(\mathrm{E}\{C(C - \mathrm{E}[C|P,Z])|P,Z\}\Big).\end{aligned}$$

Since $\mathrm{E}\{C(T - \mathrm{E}[T|P,Z])|P,Z\} = \mathrm{Cov}[C,T|P,Z]$ and $\mathrm{E}\{C(C - \mathrm{E}[C|P,Z])|P,Z\} = \mathrm{Var}[C|P,Z]$, solving for μ gives (46) when $(P,Z) = (p,z)$. Then $\lambda(z)$ is obtained from $\mu(z) = \lambda(z)/[1 + \lambda(z)]$. Moreover, $\mathrm{Cov}[C,T|P,Z] = \mathrm{E}\{C(T - \mathrm{E}[T|P,Z])|P,Z\} = -[\Gamma(P,Z)/\overline{c}_o(P,Z)]\mathrm{Var}[C|P,Z]$ by (B.16) and Assumption B4. Thus, $\mathrm{Cov}[C,T|P,Z] < 0$ as $\Gamma(\cdot,\cdot) > 0$ and $\overline{c}_o(\cdot,\cdot) > 0$.□

Proof of Proposition 8: Let $S = [y, c_o, H, \psi, G, F, \lambda] \in \mathcal{S}'$ generating the observations (Y,C,P,T) given Z. We construct a structure $S = [\tilde{y}, \tilde{c}_o, \tilde{\psi}, \tilde{G}, \tilde{F}, \tilde{\lambda}] \in \mathcal{S}$ with $H(\cdot,\cdot)$ the identity function. In view of the nonidentification of $\lambda(\cdot)$ in the basic model as shown in Proposition 6, one can choose $\tilde{\lambda}(\cdot) = \lambda(\cdot)$. We can also consider that the error terms and unobserved heterogeneity term remain the same leading to $\tilde{G}(\cdot,\cdot,\cdot|\cdot) = G(\cdot,\cdot,\cdot|\cdot)$. Moreover, let $\tilde{y}(\cdot,\cdot,\cdot) = y(\cdot,\cdot,\cdot)$ and $\tilde{c}_o(\cdot,\cdot,\cdot) = c_o(\cdot,\cdot,\cdot)$. Hence, from (25) and (26), we have $\overline{\tilde{y}}(p,z) = \overline{y}(p,z)$ and $\overline{\tilde{c}}_o(p,z) = \overline{c}_o(p,z)$. We note that $\mathrm{E}[C|P=p,Z=z] = H(\theta-e,z)\overline{c}_o(p,z) = (\tilde{\theta}-\tilde{e})\overline{\tilde{c}}_o(p,z)$. Thus, $H(\theta-e,z) = \tilde{\theta}-\tilde{e}$ or equivalently $H[\Delta(p,z),z] = \tilde{\Delta}(p,z)$. From observations (Y,C,P,T), one can define the disutility function $\tilde{\psi}(e,z)$ and the optimal effort function $\tilde{e}(p,z)$ from (33) and (34), where $\Delta(\cdot,\cdot)$, $\Gamma(\cdot,\cdot)$ and $R(\cdot,\cdot)$ are replaced by $\tilde{\Delta}(\cdot,\cdot)$, $\tilde{\Gamma}(\cdot,\cdot)$ and $\tilde{R}(\cdot,\cdot)$, respectively. The distribution $\tilde{F}(\cdot|\cdot)$ is then constructed as the distribution of the term $\tilde{\theta}$ defined in (35) with $\Delta(\cdot,\cdot)$ and $R(\cdot,\cdot)$ replaced by $\tilde{\Delta}(\cdot,\cdot)$ and $\tilde{R}(\cdot,\cdot)$, respectively.

By the extension of Lemma 6 with $H(\cdot,\cdot)$ given, the conditional distribution of (Y,C,P,T) given Z induced by the structure S must satisfy assumption D2. We first show that this distribution also satisfies assumption C1. Assumption C1-(i) is trivially satisfied. Regarding assumption C1-(ii), we have $\Gamma(p,z) = -(\partial\mathrm{E}[T|P=p,Z=z]/\partial p)/(\partial\Delta(p,z)/\partial p)$, $\tilde{\Gamma}(p,z) = -(\partial\mathrm{E}[T|P=p,Z=z]/\partial p)/(\partial\tilde{\Delta}(p,z)/\partial p)$ with $\tilde{\Delta}(p,z) = H(\Delta(p,z),z)$. Thus $\tilde{\Gamma}(p,z) = \Gamma(p,z)/H'(\Delta(p,z),z)$. Since $H'(\cdot,\cdot) > 0$ by assumption A2-(iii), we have $\tilde{\Gamma}(p,z) > 0$ as $\Gamma(p,z) > 0$ by assumption D2-(ii). Regarding its derivative, we obtain

$$\frac{\partial\tilde{\Gamma}(p,z)}{\partial p} = \frac{H'(\Delta(p,z),z)(\partial\Gamma(p,z)/\partial p) - \Gamma(p,z)H''(\Delta(p,z),z)(\partial\Delta(p,z)/\partial p)}{H'^2(\Delta(p,z),z)} < 0,$$

since $H'(\cdot,\cdot) > 0$ and $H''(\cdot,\cdot) \geq 0$ by assumption A2-(iii), and $\Gamma(p,z) > 0$, $\partial\Gamma(p,z)/\partial p < 0$ and $\partial\Delta(p,z)/\partial p > 0$ by assumption D2-(ii,iii). Thus assumption C1-(ii) is satisfied. Regarding C1-(iii), we have $\partial\tilde{\Delta}(p,z)/\partial p = H'(\Delta(p,z),z)(\partial\Delta(p,z)/\partial p) > 0$ since $H'(\cdot,\cdot) > 0$ by assumption A2-(iii) and $\partial\Delta(p,z)/\partial p > 0$ by assumption D2-(iii). Since $\tilde{\Gamma}(p,z) = \Gamma(p,z)/H'(\Delta(p,z),z)$. From assumption D2-(iii), we have $\Gamma(p,z) < H'(\Delta(p,z),z)\overline{c}_o(p,z)$. Since $H'(\cdot,\cdot) > 0$ by assumption A2-(iii), this gives $\tilde{\Gamma}(p,z) < \overline{c}_o(p,z)$. Regarding C1-(iv), we have $\epsilon_d = \phi_d(Y,P,Z) = \tilde{\phi}_d(Y,P,Z)$, $\epsilon_c = \phi_c(Y,P,C,Z) = \tilde{\phi}_c(Y,P,C,Z)$ and $\epsilon_t = \phi_t(Y,P,C,T,Z) = \tilde{\phi}_t(Y,P,C,T,Z)$ showing that assumption C1-(iv) is satisfied in view of assumption D2-(iv). Lastly, assumption C1-(v) is trivially satisisfied in view of assumption D2-(v).

We now invoke the second part of Lemma 7, which establishes the existence of a basic structure in $\mathcal{S}$ that rationalizes the conditional distribution of (Y,C,P,T) given Z. Hence, this basic structure is observationally equivalent to S. $\square$

References

Baron, D. (1989): "Design of Regulatory Mechanisms and Institutions," in R. Schlamensee and R. Willig, eds., *Handbook of Industrial Organization, Volume II*, North Holland.

Baron, D. and **R. Myerson** (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica,* 50, 911-930.

Brocas, I., **K. Chan** and **I. Perrigne** (2006): "Regulation under Asymmetric Information in Water Utilities," *American Economic Review, Papers and Proceedings*, 96, 62-66.

Chernozhukov, V., **H. Hong** and **E. Tamer** (2006): "Inference on Parameter Sets in Econometric Models," Working Paper, Northwestern University.

Chiappori P.A. and **B. Salanié** (2003): "Testing Contract Theory: A Survey of Some Recent Work," in M. Dewatripont, L. Hansen and S. Turnovsky, eds., *Advances in Economics and Econometrics, Vol. I*, Cambridge University Press.

Février, P., **R. Preget** and **M. Visser** (2004): "Econometrics of Share Auctions," Working Paper, CREST.

Gagnepain, P. and **M. Ivaldi** (2002): "Incentive Regulatory Policies: The Case of Public Transit Systems in France," *Rand Journal of Economics*, 33, 605-629.

Guerre, E., **I. Perrigne** and **Q. Vuong** (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525-574.

Guerre, E., **I. Perrigne** and **Q. Vuong** (2006): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," Working Paper, Pennsylvania State University.

Huang, Y., **I. Perrigne** and **Q. Vuong** (2007): "Nonparametric Identification and Estimation of Nonlinear Pricing Models with an Application to Yellow Pages," in progress.

Joskow, N. and **N. Rose** (1989): "The Effects of Economic Regulation," in R. Schlamensee and R. Willig, eds., *Handbook of Industrial Organization, Volume II*, North Holland.

Koopmans, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125-144.

Laffont, J.J. (1994): "The New Economics of Regulation Ten Years After," *Econometrica*, 62, 507-537.

Laffont, J.J. (2005): *Regulation and Development*, Cambridge University Press.

Laffont, J.J. and **D. Martimort** (2001): *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.

Laffont, J.J. and **J. Tirole** (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614-641.

Laffont, J.J. and **J. Tirole** (1993): *A Theory of Incentives in Procurement and Regulation*, MIT Press.

Laffont, J.J. and **Q. Vuong** (1996): "Structural Analysis of Auction Data," *American Economic review, Papers and Proceedings,* 86, 414-420.

Li, T., **I. Perrigne** and **Q. Vuong** (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics,* 98, 129-161.

Manski, C. and **E. Tamer** (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-547.

Matzkin, R. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Volume IV*, North Holland.

Matzkin, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

Matzkin, R. (2005): "Identification in Nonparametric Simultaneous Equations," Working Paper, Northwestern University.

Miravete, E. (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *Review of Economic Studies*, 69, 943-971.

Paarsch, H. (1992): "Deciding Between the Common Value and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191-215.

Park, B. and **L. Simar** (1994): "Efficient Semiparametric Estimation in Stochastic Frontier Models," *Journal of the American Statistical Association*, 89, 929-936.

Park, B., **R. Sickles** and **L. Simar** (1998): "Stochastic Panel Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273-301.

Perrigne, I. (2002): "Incentive Regulatory Contracts in Public Transportation: An Empirical Study," Working Paper, Pennsylvania State University.

Perrigne, I. and **S. Surana** (2004): "Politics and Regulation: The Case of Public Transit," Working Paper, Pennsylvania State University.

Perrigne, I. and **Q. Vuong** (2007a): "Testing in Incentive Regulation Models," in progress.

Perrigne, I. and **Q. Vuong** (2007b): "Estimating Cost Efficiency Under Asymmetric Information," in progress.

Prakasa Rao, B. (1992): *Identifiability in Stochastic Models*, Academic Press.

Roehrig, C. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433-447.

Wolak, F. (1994): "An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction," *Annales d'Economie et de Statistiques*, 34, 13-69.