A Structural Analysis of Simple Contracts*

Yonghong An† Shengjie Hong‡ Daiqiang Zhang§

April 6, 2022

Abstract

This paper provides an econometric framework for analyzing two-period simple contracts where an agent chooses between a fixed-price option and a cost-reimbursement option provided by a principal in each contracting period. First, we propose a consistent procedure for testing the null hypothesis of a corresponding cost function being linear, which is widely assumed for tractability in the literature. Motivated by the rejection of such a null based on our empirical data, next we establish nonparametric identification, without restricting the cost function to be linear, for all model primitives conditioned on the agent exerting nonzero effort. These primitives include agent's cost and disutility functions, distribution of agent efficiency type, and parameters that characterize agent's bargaining power and intertemporal preference. Moreover, we propose a consistent procedure to implement the identification results for estimation. In our empirical study, we find strong evidence against linearity of the cost function. The importance of this empirical finding is further evidenced by a welfare analysis, which shows the welfare assessment to be sensitive to the specification of cost function.

Keywords: Simple contracts, two-period, measurement error, nonparametric identification.
JEL: C14, D82.

*We would like to thank Xiaohong Chen, Qi Li, Yu-Chin Hsu, Ping Yu, and participants of various conferences and seminars for helpful comments, discussion, and suggestions. Hong gratefully acknowledges support from the National Natural Science Foundation of China (Grant No. 71703078). This paper supersedes a previous working paper "Econometrics of Multi-Period Simple Contracts". All remaining errors are ours.

†Department of Economics, Texas A&M University, College Station, TX 77843; email: y.an@tamu.edu.

‡School of Economics, Central University of Finance and Economics, Beijing, China 100081; email: hongshj@vip.163.com.

§Department of Economics, University at Albany, State University of New York, Albany, NY 12222; email: dzhang6@albany.edu.

1 Introduction

Due to the fundamental role it plays in understanding informational asymmetries and incentives, contract theory has been a rapidly developing field in economics for the past three decades. Early studies had been focused on *complex optimal contracts*, in the spirit of Laffont and Tirole (1986), where the payment scheme to an agent is specified as a function of both the agent's observed cost and his unobserved type. More recently, *simple contracts*, which are also known as (a.k.a.) *simple menu contracts* in the literature, have attracted a lot of attention. This branch of contracts specify the payment scheme simply as a function of only the agent's observed cost or even as a constant. Not surprisingly, simple contracts are more widely adopted in practice than complex optimal contracts, because their payment schemes are much easier to implement. The theoretic importance of simple contracts has been much recognized. In particular, evidences from both theory and empirical studies have shown that simple contracts can capture a substantial proportion of the surplus that complex optimal contracts would achieve (Rogerson, 2003; D'Haultfoeuille and Février, 2019). However, econometric analyses of simple contracts are still very limited.

In this paper, we study simple contracts with two basic options: fixed-price (FP) and cost-reimbursement (CR). The FP option specifies the payment to be a fixed amount regardless of the agent's realized cost. In contrast, the CR option specifies the payment to equal exactly the agent's realized cost. FP-CR menus are widely used in practice, e.g., they are commonly adopted by the U.S. Department of Defense (Rogerson, 1992). They are also used by local authorities in France to contract with operators to provide transportation service(Aaker and Myers, 1987; Gagnepain et al., 2013). Additional examples include the Indian customized software industry (Banerjee and Duflo, 2000), the U.S. Air Force engine procurement (Bajari and Tadelis, 2001), the offshore software industry (Gopal et al., 2003), among many others. Our goal is to develop a flexible framework for econometric analysis of FP-CR contracts.

First, we construct an economic model to serve as the basis for our structural econometric analyses. We take Gagnepain et al. (2013)'s economic model as our starting point and depart from there by relaxing one of their key assumptions, namely the *linear cost condition* (LCC hereafter). The LCC restricts the agent's cost of fulfilling a contract to be linear in his efficiency type (a.k.a. innate cost). Often assumed for tractability, the LCC also helps to achieve econometric identification when coupled with other simplifying assumption(s), e.g., a normal distribution for the efficiency type as in Gagnepain et al. (2013). However, its empirical validity is questionable. For instance, some empirical evidences suggest that the optimal effort to be monotonic in the efficiency type (see, e.g., Gagnepain and Ivaldi, 2002 and Abito, 2014), as opposed to being constant (i.e., being invariant in the efficiency type) which necessarily holds under the LCC. After establishing the economic model, we specify the econometric setting and focus on two major tasks: to develop a credible inference procedure to assess whether the LCC is consistent with the data, and to develop identification strategies for the model primitives without imposing the LCC.

To assess whether the LCC is consistent with the data, we propose a consistent procedure for testing a null hypothesis that is directly implied by the LCC. The usefulness of the proposed testing procedure is two-fold: (i) The modeling simplicity under the LCC is appealing and appreciated, especially if there is no significant empirical evidence against it. So if the proposed test fails to reject the null, it is acceptable to impose the LCC; (ii) If the null is rejected, it is better to adopt a more credible way of econometric analyses, which does not rely on the LCC.

The other major task is to develop identification strategies for the model primitives without imposing the LCC. These primitives include agent's cost function, agent's disutility function (from exerting cost-reducing effort), distribution of agent's efficiency type, and parameters relating to agent's bargaining power and intertemporal preference. These identifications are achieved through the following steps: First, by adopting a recent method on measurement error by Schennach and Hu (2013), we recover the distribution of the unobserved optimal effort from the joint distribution of two observable effort-related proxies; Next, we work on identification of the cost function, which is the key to identify all other model primitives. We require the existence of an exclusion variable that is independent from agent's type but affects the disutility from exerting cost-reducing efforts. Our strategy to identify the cost function is to match quantiles of the cost and effort distributions conditioned on different values of the exclusion variable, according to the corresponding quantiles of the type distribution, which are invariant to the variable; Third, with the cost function identified, we exploit the structural link between the agent's efficiency type and the corresponding optimal effort level to recover the type value associated with any given observed cost under FP contracts; Last, based on what have been identified, we identify the remaining primitives. Following the identification strategies, we propose consistent methods to estimate all model primitives. Throughout our analysis, we prioritize on identification. To the best of our knowledge, ours is the first set of positive results on identification of multi-period simple contracts in the literature.

As an empirical illustration, we apply our methods to study transportation procurement contracts in France. The main objectives are to test the widely assumed LCC, and to evaluate how this specification assumption affects welfare assessment. Based on our empirical data, a direct implication from the LCC is reject at a significance level of 1%. This testing result suggests that the LCC fails to adequately describe the cost structure in the French transportation industry. To evaluate how the LCC affects the welfare assessment, we estimate the welfare with and without imposing the LCC then conduct a comparison. We find substantial difference between the two assessments, which suggests the welfare assessment to be sensitive to the specification of cost function. Thus, one needs to be cautious on deciding whether or not to impose linearity, since mis-specification could lead to substantial bias.

Our study contributes to the growing literature on identification of contract models; see, Perrigne and Vuong (2011) and D'Haultfoeuille and Février (2019), among others. Perrigne and

Vuong (2011) establish nonparametric identification of a static *complex* contract model tailored from the seminal paper Laffont and Tirole (1986). D'Haultfoeuille and Février (2019) show nonparametric *partial identification* of simple contracts using exogenous variations of contracts. We note that the identification strategy of Perrigne and Vuong (2011) is not applicable to our setting, mainly because the one-to-one mapping between observed product price and agent type, a key to achieve identification in Perrigne and Vuong (2011), is unavailable from the simple contracts that we consider. Compared with D'Haultfoeuille and Février (2019), our paper differs in both model setup and identification strategies.[1] Our paper is closely related to Gagnepain et al. (2013) in the sense that our economic model of FP-CR contracts adopts a similar principal-agent setup. Besides, we use the same dataset of French public transportation service contracts as they do for our empirical study. Nevertheless, Gagnepain et al. (2013) focus on evaluating the social cost of allowing for renegotiation, where the whole analysis is under the assumption that the LCC holds. While we mainly focus on testing the validity of the LCC and developing identification results without imposing it.

In a broader view, our identification result is related to the econometric analysis on a richer class of principal-agent models featuring moral hazard and adverse selection, where an agent with unknown type is offered a menu of only a few simple contractual options by the principal, such as the insurance models studied by Aryal et al. (2010) and the nonlinear pricing models studied by Luo et al. (2018a). In addition, a growing number of papers utilize developments in measurement errors to identify structural models. See Hu (2017) for a recent survey. Our paper also contributes to this category of literature: As far as we know, our study is the first to employ the results in measurement errors to identify contract models.

The rest of the paper is organized as follows. In Section 2 we present the economic model of FP-CR contracts. In Section 3 we specify our econometric setting and propose a formal test for the LCC. In Section 4 we establish the main identification results. In Section 5 we propose consistent estimation procedures. In Section 6 we provide some Monte Carlo evidence of our method. In Section 7 we apply our method to French transportation procurement contracts. We conclude the paper in Section 8. Proofs of theorems, lemmas and corollaries are collected in the Appendix.

2 The model

A local authority (the 'principal') wants to procure a public service or project from a firm (the 'agent'). A procurement contract can last for a single period or for multiple periods. A single-period contract sets a payment option, with all associated terms fully specified, from

[1]In D'Haultfoeuille and Février (2019), unlike standard contract models, agent's disutility function of effort is not explicitly modeled. In contrast, we model the disutility function explicitly. Moreover, while D'Haultfoeuille and Février (2019) provide very informative partial identification results, we aim at developing regularity conditions for point identification.

a menu of options offered by the principal. A multi-period contract sets a sequence of fully specified options, one for each period. Let $T < \infty$ be the total number of contracting periods. A multi-period contract can be signed with *full commitment* or with *renegotiation* permitted. In the latter case, there is a window in between periods t and $t+1$ (i.e., between the end of period t and the beginning of period $t+1$), for each $t = 1, ..., T-1$, during which the principal and agent are allowed to renegotiate to alter options and/or their detailed terms for subsequent periods (i.e., from $t+1$ to T). In contrast, a contract with full commitment prohibits any renegotiation once signed.

Throughout the paper, we primarily focus on two-period contracts (i.e., $T = 2$) that permit renegotiation. Our proposed methods for econometric analyses work for single-period contracts (i.e., $T = 1$) with only slight modifications needed, as explained in Appendix D.2. The proposed methods, for the most part, can be straightforwardly extended to $T > 2$ periods, with some complication arising from equilibrium characterization, which we discuss at the end of Section 4 after establishing our main identification results.

2.1 Basic setup

The agent's per period cost c_t is determined by his efficiency type θ (a.k.a. innate cost) and his cost reducing effort e_t, according to a general cost function $H(\cdot)$ as

$$c_t = H(\theta - e_t), \quad \text{for } t = 1, ..., T.$$

Often interpreted in the literature as a measure of management and production skills, θ is randomly drawn from a cumulative distribution function (CDF) $F(\cdot)$, with density $f(\cdot)$, on a bounded support $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$. θ is the agent's private information and is specified as time-invariant, while $F(\cdot)$ is common knowledge. $e_t \geq 0$ measures the agent's cost-reducing effort in period t, which induces some disutility according to a disutility function $\psi(e_t)$. We impose the following conditions, which are standard in related literature (e.g., Laffont and Tirole, 1993):

Assumption 2.1 *(i) $H(\cdot) \geq 0$, $H'(\cdot) > 0$ and $H''(\cdot) > 0$; (ii) $\psi(\cdot) \geq 0$, $\psi'(\cdot) > 0$, $\psi''(\cdot) > 0$ and $\psi(0) = 0$.*

The agent's payoff in period t is

$$q_t - c_t - \psi(e_t), \text{ with } q_t = \begin{cases} c_t \text{ under CR;} \\ b_t \text{ under FP,} \end{cases}$$

where q_t is the payment he receives from the principal, and b_t is the fixed price associated with the FP option. Given b_t, the agent chooses between FP and CR, and decides an optimal level of effort, to maximize his payoff. Under FP, the optimal level of effort depends on θ in general, hence denoted by $e(\theta)$. We have

$$e(\theta) = \underset{e}{\operatorname{argmin}} \{H(\theta - e) + \psi(e)\}, \tag{2.1}$$

which satisfies the first-order condition (f.o.c.)

$$H'(\theta - e(\theta)) = \psi'(e(\theta)). \tag{2.2}$$

Consequently, under FP with a fixed price b_t, the maximum payoff achievable is

$$u_F(\theta, b_t) \equiv b_t - H(\theta - e(\theta)) - \psi(e(\theta)). \tag{2.3}$$

Since the CR option provides no incentive for any cost reducing effort, the optimal effort level under CR would be zero regardless of θ. As a result, under the normalizing condition $\psi(0) = 0$ in Assumption 2.1, the payoff under CR would always be zero. So the agent will choose FP (with b_t) over CR in period t if and only if $u_F(\theta, b_t) \geq 0$. With optimal choices of option and effort level, his per period payoff would be

$$u(\theta, b_t) \equiv \max\{u_F(\theta, b_t), 0\}. \tag{2.4}$$

The principal is primarily concerned with social welfare from fulfillment of the contract. In generic terms, the social welfare is defined as

$$SW = S - (1+\lambda) Q + \alpha U \tag{2.5}$$

where S is the gross surplus generated by the procured service, Q is the principal's payment to the agent, and U is the agent's payoff. The way the principal raises the payment funds Q, usually by imposing distortion tax, typically leads to a dead-weight loss. Such a dead-weight loss is captured by λ. α measures the agent's bargaining power. From an alternative view, α can be interpret as reflecting the local authority's political preferences. The definition (2.5) above dates back to Baron and Myerson (1982), and have been widely used in the contract literature.[2]

Suppose, in period t, the principal makes a payment q_t to the agent, and the agent's payoff ends up being u_t. According to the general definition (2.5), the per period social welfare is

$$\pi_t = S - (1+\lambda) q_t + \alpha u_t,$$

Anticipating the agent's choices of option and effort level in response to a fixed price offering of b_t, the principal sets b_t to maximize expected social welfare, which we explain in details as we discuss equilibrium properties in Sections 2.2 and 2.3. Before we proceed, we note that the following per period patterns hold in both the single-period and two-period settings:

Lemma 2.1 *Let Assumption 2.1 hold. (i) Under CR, the agent exerts zero effort regardless of θ, and the cost is realized as $c_t = H(\theta)$; (ii) Under FP, the optimal effort $e(\theta)$ is strictly increasing in θ, with $e'(\theta) \in (0,1)$; (iii) Moreover, under FP, both $\theta - e(\theta)$ and $H(\theta - e(\theta))$ (i.e., the realized cost c_t) are strictly increasing in θ.*

[2] For more discussions on λ and α, see Baron and Myerson, 1982, Baron, 1988 and Gagnepain et al., 2013. Also, S is assumed to be sufficiently large to guarantee the desirability of the procured service or project.

2.2 Equilibrium for single-period contracts

It is helpful to examine the single-period setting first, for understanding two-period settings. For simplicity of notation, we drop the subscript t (that indicates the time period) for the single-period setting. In equilibrium, for any fixed price b she may offer, the principal correctly anticipates the agent's choices of option and effort level in response conditional on θ, as specified by (2.1), as well as his resulting payoff, as specified by (2.4). She predicts the social welfare resulting from offering b, conditional on θ, to be

$$\tilde{\pi}(\theta, b) = S - (1+\lambda)\left\{\mathbf{1}\left(u_F(\theta, b) \geq 0\right) \cdot b + \left[1 - \mathbf{1}\left(u_F(\theta, b) \geq 0\right)\right] \cdot H(\theta)\right\} + \alpha u(\theta, b). \quad (2.6)$$

So she sets an optimal fixed price b^* to maximize the *expected social welfare* defined as

$$\pi(b) \equiv \mathbb{E}_\theta\left[\tilde{\pi}(\theta, b)\right] = \int_\Theta \tilde{\pi}(\theta, b)\, dF(\theta). \quad (2.7)$$

Proposition 2.1 (Single-period equilibrium) *Consider a single-period contract. Let Assumption 2.1 hold. In equilibrium, the following holds:*
(i) There is a unique cutoff value $\theta^* \in \left[\underline{\theta}, \bar{\theta}\right]$ *such that the agent chooses FP if* $\theta \leq \theta^*$*, i.e., being more efficient, and chooses CR otherwise.*
(ii) The principal offers the optimal fixed price $b^* = \underset{b}{\operatorname{argmax}}\, \pi(b)$*.* b^* *and* θ^* *satisfy*

$$b^* = H\left(\theta^* - e\left(\theta^*\right)\right) + \psi\left(e\left(\theta^*\right)\right), \quad (2.8)$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right) \frac{F(\theta^*)}{f(\theta^*)} = \frac{H(\theta^*) - b^*}{H'\left(\theta^* - e\left(\theta^*\right)\right)}. \quad (2.9)$$

2.3 Equilibrium for two-period contracts

For a two-period FP-CR contract, the principal can offer four possible combinations of per period options. The option-menu is as follows.

$$\begin{cases} \textbf{FF}\text{: FP for both periods, with payments } \left(b_1^{FF}, b_2^{FF}\right); \\ \textbf{CF}\text{: CR for period-1 and FP for period-2, with payments } \left(c_1, b_2^{CF}\right); \\ \textbf{CC}\text{: CR for both periods, with payments } \left(c_1, c_2\right); \\ \textbf{FC}\text{: FP for period-1 and CR for period-2, with payments } \left(b_1^{FC}, c_2\right). \end{cases}$$

We ignore the fourth option, FC, as it would never appear in equilibrium.[3]

Use generic notations u_t and π_t for the payoff and social welfare in period t, respectively, for $t = 1, 2$. The *intertemporal payoff* and *intertemporal social welfare* are defined respectively as

$$\gamma u_1 + (1-\gamma)\, u_2 \quad \text{and} \quad \gamma \pi_1 + (1-\gamma)\, \pi_2,$$

[3] This is because, when offered a vector of fixed prices $\left(b_1^{FF}, b_2^{FF}, b_2^{CF}, b_1^{FC}\right)$ that maximizes the principal's objective function (to be specified soon for the two-period setting), the agent would always prefer one of the first three options (i.e., FF, CF and CC) to FC regardless of θ.

where $\gamma \equiv 1/(1+\delta)$ (with δ being the discount factor) is a measure of the relative importance of the first period, following Laffont and Tirole (1990)'s interpretation.

Given a vector of fixed prices $\mathbf{b} = \left(b_1^{FF}, b_2^{FF}, b_2^{CF}\right)'$, the maximum agent's payoffs under FF, CF and CC are

$$u_{FF}(\theta, \mathbf{b}) \equiv \gamma u_F\left(\theta, b_1^{FF}\right) + (1-\gamma)\, u_F\left(\theta, b_2^{FF}\right),$$
$$u_{CF}(\theta, \mathbf{b}) \equiv (1-\gamma)\, u_F\left(\theta, b_2^{CF}\right) \text{ and } u_{CC}(\theta, \mathbf{b}) \equiv 0,$$

respectively, where $u_F(\theta, \cdot)$ is defined in (2.3). These maximums are achieved by making an effort of level $e(\theta)$ (as specified in (2.1)) for any FP period, and zero effort for any CR period. Based on θ and $\mathbf{b}$, the agent would choose option $J \in \{FF, CF, CC\}$ that yields the largest $u_J(\theta, \mathbf{b})$ to maximize his intertemporal payoffs. So the agent's choice of option, denoted by $J(\theta, \mathbf{b})$, is characterized as

$$J(\theta, \mathbf{b}) = \underset{J \in \{FF, CF, CC\}}{\operatorname{argmax}} \; u_J(\theta, \mathbf{b}).$$

In equilibrium, correctly anticipating the agent's choices of option and effort level as described above, the principal predicts the intertemporal social welfare conditional on θ to be

$$\tilde{\pi}_{int}(\theta, \mathbf{b}) = S - (1+\lambda)\, q_{J(\theta,\mathbf{b})}(\theta, \mathbf{b}) + \alpha u_{J(\theta,\mathbf{b})}(\theta, \mathbf{b}) \tag{2.10}$$

for any given vector of fixed prices $\mathbf{b} = \left(b_1^{FF}, b_2^{FF}, b_2^{CF}\right)'$, where

$$q_J(\theta, \mathbf{b}) \equiv \begin{cases} \gamma b_1^{FF} + (1-\gamma)\, b_2^{FF}, & \text{under FF;} \\ \gamma H(\theta) + (1-\gamma)\, b_2^{CF}, & \text{under CF;} \\ H(\theta), & \text{under CC} \end{cases} \tag{2.11}$$

is the intertemporal payment she needs to raise under option J, for $J \in \{FF, CF, CC\}$. In equilibrium, the principal sets an optimal vector of fixed prices $\mathbf{b}^*$ to maximize the *expected intertemporal social welfare* defined as

$$\pi_{int}(\mathbf{b}) \equiv \mathbb{E}_\theta\left[\tilde{\pi}_{int}(\theta, \mathbf{b})\right] = \int_\Theta \tilde{\pi}_{int}(\theta, \mathbf{b})\, dF(\theta), \tag{2.12}$$

i.e., $\mathbf{b}^* = \underset{\mathbf{b}}{\operatorname{argmax}}\, \pi_{int}(\mathbf{b})$.

As mentioned earlier, we distinguish between two types of two-period contracts: (i) Contracts with *full commitment*; (ii) Contracts permitting *renegotiation*. Our focus is on the latter, which also fits our empirical study. Nevertheless, we first discuss the former briefly, for a better understanding of the latter.

Full commitment

A contract with full commitment prohibits any adjustment of initial arrangement during implementation. As pointed out by Laffont and Tirole (1990), the two-period optimal contract with full commitment is simply a twice-repeated version of the single-period one. Hence, it holds, for any optimal vector of fixed prices $\mathbf{b}^* = \left(b_1^{FF*}, b_2^{FF*}, b_2^{CF*}\right)' \in \underset{\mathbf{b}}{\operatorname{argmax}}\, \pi_{int}(\mathbf{b})$, that

$$b_2^{CF*} \leq b_1^{FF*} = b_2^{FF*} = b^* = \underset{b}{\operatorname{argmax}}\, \pi(b)\,. \tag{2.13}$$

There is a unique cutoff value $\theta^* \in \left[\underline{\theta}, \bar{\theta}\right]$ such that the agent chooses FF if $\theta \leq \theta^*$, and CC otherwise.[4] For b^* and θ^*, (2.8) and (2.9) in Proposition 2.1 still hold.

Renegotiation

Being observable to the principal, the agent's choice of option and realized cost in the first contracting period reveal partial information on θ. The principal may utilize this information to update her belief on the distribution of θ, which may trigger potential renegotiation in between periods in executing a multi-period contract. Here, we adopt the *limited updating* specification as in Gagnepain et al. (2013): The principal's updating belief is based on information revealed by only the agent's choice of option, but not the realized cost. The principal's initial belief is $F(\cdot)$ with support $\left[\underline{\theta}, \bar{\theta}\right]$. For a given fixed price b, let $\theta(b)$ be the corresponding cutoff type such that an agent of type $\theta(b)$ would be indifferent between FP and CR, i.e.,

$$b = H(\theta(b) - e(\theta(b))) + \psi(e(\theta(b)))\,.$$

(It follows immediately from (2.8) that $\theta^* = \theta(b^*)$.) The agent's choosing FP (with b) over CR reveals that $\theta \leq \theta(b)$, and consequently leads to an updated belief of the principal as follows

$$\underline{F}_b(\cdot) \equiv F(\cdot)/F(\theta(b))\,, \text{ with support } [\underline{\theta}, \theta(b)]\,,$$

which is $F(\cdot)$ truncated on $[\underline{\theta}, \theta(b)]$. Based on $\underline{F}_b(\cdot)$, the principal reformulates her expected per period social welfare as $\underline{\pi}_b(\cdot) \equiv \int_\Theta \tilde{\pi}(\theta, \cdot)\, \underline{F}_b(\theta)$. Similarly, choosing CR over FP (with b) leads to an updated belief of

$$\bar{F}_b(\cdot) \equiv \frac{F(\cdot) - F(\theta(b))}{1 - F(\theta(b))}, \text{ with support } \left[\theta(b), \bar{\theta}\right],$$

which is $F(\cdot)$ being truncated on $\left[\theta(b), \bar{\theta}\right]$. Based on $\bar{F}_b(\cdot)$, the principal reformulates her expected per period social welfare as $\bar{\pi}_b(\cdot) \equiv \int_\Theta \tilde{\pi}(\theta, \cdot)\, \bar{F}_b(\theta)$. As long as $\theta(b) = \left(\underline{\theta}, \bar{\theta}\right)$, it can be shown that $\underline{\pi}_b(\cdot)$ is maximized at a value smaller than b^*, and that $\bar{\pi}_b(\cdot)$ is maximized at a value larger than b^*,

[4]Note that, with full commitment, CF would not be selected in equilibrium.

To further illustrate how renegotiation may happen, suppose the principal and agent sign an initial contract with fixed prices $\mathbf{b}^*$ as in (2.13), which would be the equilibrium profile with full commitment. Consider the following cases: (i) Suppose the agent chooses FF. Realizing that the agent is of more efficiency type (whose θ is distributed according to $\underline{F}_{b^*}(\cdot)$), the principal would want to renegotiate for a lower second period fixed price $b_2^{FF} < b_2^{FF*} = b^*$. However, the agent would not agree on a lower fixed price, so renegotiation would not actually happen in this case; (ii) Now suppose the agent chooses CC. Realizing that the agent is of less efficiency type (distributed according to $\bar{F}_{b^*}(\cdot)$), the principal would want to renegotiate for a higher second period fixed price $b_2^{CF} > b^* \geq b_2^{CF*}$ to provide more incentive for the agent to select FP and revamp efficiency in period-2. Since a higher fixed price is always welcomed by the agent, a renegotiated second-period fixed price can be possibly agreed on in this case.

According to the *renegotiation-proof principle*, any agreement of $\left\{b_1^{FF}, b_2^{FF}, b_2^{CF}\right\}$, of which the period-2 continuation $\left\{b_2^{FF}, b_2^{CF}\right\}$ is superseded by the renegotiated $\left\{\tilde{b}_2^{FF}, \tilde{b}_2^{CF}\right\}$, could be replaced by the agreement of $\left\{b_1^{FF}, \tilde{b}_2^{FF}, \tilde{b}_2^{CF}\right\}$ under which no renegotiation would actually happen. For this reason, we focus on *renegotiation-proof profiles*, which is a standard arrangement in the contract literature and causes no loss of generality. Proposition 2.2 below characterizes the equilibrium.

Proposition 2.2 (Two-period equilibrium) *Consider a two-period contract that permits renegotiation. Let Assumption 2.1 hold. In equilibrium, the following holds:*
(i) There are two cutoff values $\theta_1^*, \theta_2^* \in \left[\underline{\theta}, \bar{\theta}\right]$ with $\theta_1^* < \theta_2^*$ such that the agent chooses FF if $\theta \leq \theta_1^*$, chooses CF if $\theta_1^* < \theta \leq \theta_2^*$, and chooses CC otherwise.
(ii) The optimal renegotiation-proof fixed prices $\mathbf{b}^\dagger = \left(b_1^{FF\dagger}, b_2^{FF\dagger}, b_2^{CF\dagger}\right)'$ are such that $\underline{b} \equiv b_1^{FF\dagger} = b_2^{FF\dagger} < b_2^{CF\dagger} \equiv \bar{b}$. Moreover, $\underline{b}$, $\bar{b}$, θ_1^* and θ_2^* satisfy

$$\bar{b} = H\left(\theta_2^* - e(\theta_2^*)\right) + \psi\left(e(\theta_2^*)\right), \tag{2.14}$$

$$\underline{b} = \gamma\left[H\left(\theta_1^* - e\left(\theta_1^*\right)\right) + \psi\left(e\left(\theta_1^*\right)\right)\right] + (1-\gamma)\,\bar{b}, \tag{2.15}$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{F\left(\theta_2^*\right) - F\left(\theta_1^*\right)}{f\left(\theta_2^*\right)} = \frac{H\left(\theta_2^*\right) - \bar{b}}{H'\left(\theta_2^* - e(\theta_2^*)\right)}. \tag{2.16}$$

Proposition 2.2 above encompasses the results of Proposition 2 in Gagnepain et al. (2013). For the special case of LCC, (2.16) reduces to their Equation (4). Similar to theirs, the intuition for $\underline{b} < \bar{b}$ can be ascribed to that, under renegotiation, fixed prices must be raised sufficiently to induce agents with intermediate efficiency who prefer CR in period-1 to choose FP in period-2.

3 The econometric setting and a test for LCC

For econometric analysis of FP-CR contracts, we keep our focus on the single-period and two-period settings. We aim at two major tasks: to test for the LCC, and to develop identification strategies without the LCC. We prioritize to elaborate methods for the two-period setting in the main text, and note the followings: (i) Our prioritization on two-period helps to keep the notations and overall presentation tidy; (ii) We show in Appendix D.2 that the proposed methods work for the single-period setting with only slight modifications (and a set of simplified notations) needed.

In this section, we first specify the econometric setting for the rest of the paper. Then we propose a testing procedure for a testable implication of the LCC. In what follows we use upper-case letters for random variables, and lower-case letters for their realizations.

3.1 The econometric setting

From a two-period contract, we, as researchers, observe the per period cost C_t and payment Q_t (for $t = 1, 2$) and dummy variables D^{FF} and D^{CF} that indicate the choice among FF, CF and CC. For $J \in \{FF, CF\}$, $D^J = 1$ indicates J being chosen. $D^{FF} = D^{CF} = 0$ indicates CC. In addition, we require a binary variable $W \in \{\varpi_1, \varpi_2\}$ that satisfies the following condition, which serves as an exclusion variable:

Condition T.1 *W is independent of θ.*

In summary, we observe a random sample collected from n two-period contracts:

$$\left\{c_{it}, q_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1,\ldots,n;\, t=1,2}.$$

As to be shown, the testing procedure (for the LCC) can be based on either the period-1 sample $\left\{c_{i1}, q_{i1}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1}^n$ or the period-2 one $\left\{c_{i2}, q_{i2}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1}^n$, so are most of the identification steps. Both the testing procedure and identification steps heavily utilize the following period-t subsample, define, for either $t = 1$ or 2, as

$$\left\{c_{it}, q_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i\in\mathcal{FP}_t} \text{ with } \mathcal{FP}_t = \begin{cases} \left\{i : d_i^{FF} = 1\right\} \text{ for } t = 1; \\ \left\{i : d_i^{FF} + d_i^{CF} = 1\right\} \text{ for } t = 2, \end{cases} \tag{3.1}$$

i.e., the subsample of contracts with FP in period-t,[5] which can be further divided into two groups by conditioning on W:

$$\mathcal{FP}_t^j \equiv \{i \in \mathcal{FP}_t : w_i = \varpi_j\}, \text{ for } j = 1, 2. \tag{3.2}$$

A few additional notations are needed, most of which are conditional (on W) version of ones already introduced, as follows: $\psi_j(\cdot) \equiv \psi(\cdot; \varpi_j)$ denotes the disutility function associated with $W = \varpi_j$, and $\psi_j'(e) \equiv \partial\psi(e; \varpi_j)/\partial e$ denotes its derivative; $e(\theta; \varpi_j)$ denotes the optimal effort under FP given θ, conditional on $W = \varpi_j$, which is characterized by the f.o.c.

$$H'(\theta - e(\theta; \varpi_j)) = \psi_j'(e(\theta; \varpi_j)); \tag{3.3}$$

[5]The subsample with FP in period-1 consists of only FF contracts. The subsample with FP in period-2 consists of FF and CF contracts, but no CC ones.

$c(\theta;\varpi_j) \equiv H(\theta - e(\theta;\varpi_j))$ denotes the cost under the optimal effort under FP, conditional on $W=\varpi_j$; For any given $\tau \in [0,1]$, $\theta(\tau)$ denotes the τ'th quantile of $F(\cdot)$, i.e., the unconditional distribution of θ; θ_t^{j*} denotes the the cutoff value of θ between FP and CR in period t conditional on $W=\varpi_j$, for $j=1,2$. (For more details on θ_t^{j*}, please refer to Corollary 3.1 (ii)); $E_t(\tau;\varpi_j)$ denotes the τ'th quantile of an effort distribution generated by the transformation $e(\cdot,\varpi_j)$ of the truncated distribution of θ on $\left[\underline{\theta}, \theta_t^{j*}\right]$. In other words, $E_t(\tau;\varpi_j)$ is the τ'th conditional quantile of period-t effort under FP on $W=\varpi_j$.; Similarly, $C_t(\tau;\varpi_j)$ denotes the τ'th conditional quantile of period-t cost under FP on $W=\varpi_j$.

Regarding the observed variables and their samples, we impose a set of regularity conditions throughout the econometric analysis, collected in Assumption 3.1 below.

Assumption 3.1 *The following conditions hold: (i)* $\left\{c_{it}, q_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1,\ldots,n;\, t=1,2}$ *are independent and identically distributed across i for each n; (ii) The density function* $f(\cdot)$ *of* θ *exists, and is bounded and continuous on a bounded support; (iii) The conditional density functions* $f_{C_t|W}(\cdot|\varpi_j)$ *of period-t cost on* $W=\varpi_j$ *(for* $t=1,2$ *and* $j=1,2$*) under FP exist, and are bounded and uniformly continuous on bounded supports.*

And we modify conditions in Assumption 2.1 as follows to suit the econometrics analysis:

Assumption 3.2 *(i)* $H(\cdot) \geq 0$, $H'(\cdot) > 0$ *and* $H''(\cdot) > 0$*; (ii)* $\psi_j(\cdot) \geq 0$, $\psi_j'(\cdot) > 0$, $\psi_j''(\cdot) > 0$ *and* $\psi_j(0) = 0$*, for* $j=1,2$*.*

Under Assumption 3.2, all results in Lemma 2.1, Propositions 2.1 and 2.2 hold conditioned on W. Being perhaps repetitive, we (re)state these results below for convenience of reference later in the paper.

Corollary 3.1 *Let Assumption 3.2 hold. Conditioned on* $W=\varpi_j$*, it holds in equilibrium that:*
(i) Under FP, the optimal effort $e(\theta;\varpi_j)$ *is strictly increasing in* θ*, with* $\partial e(\theta;\varpi_j)/\partial\theta \in (0,1)$*. Moreover, under FP, both* $\theta - e(\theta;\varpi_j)$ *and* $H(\theta - e(\theta;\varpi_j))$ *are strictly increasing in* θ*.*
(ii) For the two-period setting, there are two cutoff values $\theta_1^{j*}, \theta_2^{j*} \in \left[\underline{\theta}, \bar{\theta}\right]$ *with* $\theta_1^{j*} < \theta_2^{j*}$ *such that the agent chooses FF if* $\theta \leq \theta_1^{j*}$*, chooses CF if* $\theta_1^{j*} < \theta \leq \theta_2^{j*}$*, and chooses CC otherwise. The optimal renegotiation-proof fixed prices* $\mathbf{b}^{j\dagger} = \left(b_1^{j,FF\dagger}, b_2^{j,FF\dagger}, b_2^{j,CF\dagger}\right)'$ *are such that* $\underline{b}^j \equiv b_1^{j,FF\dagger} = b_2^{j,FF\dagger} < b_2^{j,CF\dagger} \equiv \bar{b}^j$*. Moreover,* $\underline{b}^j$*,* $\bar{b}^j$*,* θ_1^{j*} *and* θ_2^{j*} *satisfy*

$$\bar{b}^j = H\left(\theta_2^{j*} - e\left(\theta_2^{j*};\varpi_j\right)\right) + \psi_j\left(e\left(\theta_2^{j*};\varpi_j\right)\right), \tag{3.4}$$

$$\underline{b}^j = \gamma\left[H\left(\theta_1^{j*} - e\left(\theta_1^{j*};\varpi_j\right)\right) + \psi_j\left(e\left(\theta_1^{j*};\varpi_j\right)\right)\right] + (1-\gamma)\,\bar{b}^j, \tag{3.5}$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{F\left(\theta_2^{j*}\right) - F\left(\theta_1^{j*}\right)}{f\left(\theta_2^{j*}\right)} = \frac{H\left(\theta_2^{j*}\right) - \bar{b}^j}{H'\left(\theta_2^{j*} - e\left(\theta_2^{j*};\varpi_j\right)\right)}. \tag{3.6}$$

Clearly, Corollary 3.1 (i), (ii) parallel Lemma 2.1 and Proposition 2.2, respectively.

We now establish relationship between certain quantiles, which leads to a testable implication under the LCC, and also turns out to be the key to identify the cost function without the LCC. According to the definitions of $e(\cdot;\varpi_j)$ and $c(\cdot;\varpi_j)$, it holds under FP that $c(\theta;\varpi_j) = H(\theta - e(\theta;\varpi_j))$ for all θ and $j = 1,2$, which immediately implies that $\theta = H^{-1}(c(\theta;\varpi_j)) + e(\theta;\varpi_j)$ for all θ (Note that the invertibility of $H(\cdot)$ is guaranteed by Assumption 3.2 (i)). Consequently, we have

$$\theta(\tau) = H^{-1}(c(\theta(\tau);\varpi_j)) + e(\theta(\tau);\varpi_j), \forall \tau \in [0,1]. \tag{3.7}$$

Define $p_{t,j} \equiv F\left(\theta_t^{j*}\right)$, for $t = 1,2$ and $j = 1,2$, i.e., the proportion of FP in period-t conditioned on $W = \varpi_j$. In period-t, the τ'th quantile of θ corresponds to the $(\tau/p_{t,j})$'th quantile of the truncated distribution of θ on $\left[\underline{\theta}, \theta_t^{j*}\right]$ which in turn corresponds to $C_t(\tau/p_{t,j};\varpi_j)$ and $E_t(\tau/p_{t,j};\varpi_j)$ (i.e., the $(\tau/p_{t,j})$'th conditional quantiles of period-t cost and effort, respectively, under FP on $W = \varpi_j$), for any given $\tau \in [0, p_{t,j}]$ and for $j = 1,2$. Consequently, we can rewrite (3.7) in terms of these quantiles as

$$\theta(\tau) = H^{-1}(C_t(\tau/p_{t,j};\varpi_j)) + E_t(\tau/p_{t,j};\varpi_j), \ \forall \tau \in [0, p_{t,j}]. \tag{3.8}$$

3.2 Testing for the LCC

The LCC restricts the cost function to take the form $H(\theta - e) = \beta(\theta - e)$ for some constant coefficient β. To obtain a testable implication of the LCC, we start by noting that, under the LCC, the f.o.c. (3.3) becomes $\psi_j'(e(\theta;\varpi_j)) \equiv \beta$. This together with Assumption 3.2(ii) imply that $e(\theta;\varpi_j) \equiv a_j$ for some positive constant a_j, for all $\theta \in \left[\underline{\theta}, \theta_t^{j*}\right]$, $t = 1,2$ and $j = 1,2$. [6] Consequently, under the LCC, Equation (3.8) becomes $\theta(\tau) = H^{-1}(C_t(\tau/p_{t,j};\varpi_j)) + a_j$, which in turn implies

$$C_t(\tau;\varpi_1) - C_t(\tau p;\varpi_2) \equiv \text{a constant} \, (= \beta(a_2 - a_1)), \ \forall \tau \in [0,1], \tag{3.9}$$

with $p \equiv p_{t,1}/p_{t,2}$, supposing without loss of generality that $p_{t,1} \le p_{t,2}$. Next, we show that (3.9) is testable. Specifically, for either $t = 1$ or 2 (fixed), we show how to test

$$\begin{aligned} \mathbb{H}_0: &\quad C_t(\tau;\varpi_1) - C_t(\tau p;\varpi_2) \equiv \text{a constant, for } \tau \in [0,1] \text{ almost sure}; \\ \mathbb{H}_1: &\quad \text{Otherwise.} \end{aligned}$$

The Test Statistic

Based on the subsample $\mathcal{FP}_t$ (specified in (3.1)), we construct the following *Cramér-von Mises* type test statistic

$$T_n \equiv \int_0^1 n_f \left\{ \left[\widehat{C}_t(\tau;\varpi_1) - \widehat{C}_t(\tau\hat{p};\varpi_2)\right] - \left[\widehat{C}_t(\nu;\varpi_1) - \widehat{C}_t(\nu\hat{p};\varpi_2)\right] \right\}^2 d\tau \tag{3.10}$$

[6] Since $\psi_1'(\cdot)$ and $\psi_2'(\cdot)$ may differ, it is possible that $a_1 \neq a_2$.

where: (i) $\nu \in (0,1)$ is a pre-selected constant that serves as the reference quantile level; (ii) n_f is the sample size of $\mathcal{FP}_t$, i.e., $n_f = \sum_{i=1}^{n} d_i^{FF}$ for $t=1$, $n_f = \sum_{i=1}^{n}\left(d_i^{FF} + d_i^{CF}\right)$ for $t=2$; (iii) $\hat{p} = \hat{p}_{t,1}/\hat{p}_{t,2}$ is an estimator for $p = p_{t,1}/p_{t,2}$, with

$$\hat{p}_{t,j} = \frac{\sum_{i=1}^{n} \mathbf{1}\left(i \in \mathcal{FP}_t\right) \mathbf{1}\left(w_i = \varpi_j\right)}{\sum_{i=1}^{n} \mathbf{1}\left(w_i = \varpi_j\right)}, \text{ for } j = 1, 2; \tag{3.11}$$

(iv) $\left\{\widehat{C}_t\left(\tau, \varpi_1\right), \widehat{C}_t\left(\tau\hat{p}, \varpi_2\right)\right\}$ are obtained from quantile regressions based on the period-t subsample $\left\{c_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i \in \mathcal{FP}_t}$. Specifically, $\widehat{C}_t\left(\tau, \varpi_1\right) = \widehat{q}_1\left(\tau\right)$ and $\widehat{C}_t\left(\tau\hat{p}; \varpi_2\right) = \widehat{q}_1\left(\tau\hat{p}\right) + \widehat{q}_2\left(\tau\hat{p}\right)$, where

$$\begin{aligned}
\left\{\widehat{q}_1\left(\tau\right), \widehat{q}_2\left(\tau\right)\right\} &= \underset{q_1,\, q_2}{\operatorname{argmin}} \sum_{i \in \mathcal{FP}_t} \rho_\tau\left(c_{it} - q_1 - q_2 \cdot \mathbf{1}\left(w_i = \varpi_2\right)\right), \\
\left\{\widehat{q}_1\left(\tau\hat{p}\right), \widehat{q}_2\left(\tau\hat{p}\right)\right\} &= \underset{q_1,\, q_2}{\operatorname{argmin}} \sum_{i \in \mathcal{FP}_t} \rho_{\tau\hat{p}}\left(c_{it} - q_1 - q_2 \cdot \mathbf{1}\left(w_i = \varpi_2\right)\right),
\end{aligned}$$

with $\rho_a\left(b\right) \equiv b\left[a - \mathbf{1}\left(b < 0\right)\right]$ for any given $\{a, b\}$.

Based on a uniform weak convergence result regarding a corresponding *quantile regression process* (indexed by τ) by Angrist et al. (2006) (detailed Lemma A.3 in Appendix A and its proof), we establish the asymptotic behavior of T_n, as follows

Theorem 1 (Asymptotic behavior of T_n) *Let Assumptions 3.1, 3.2, and Condition T.1 hold.*
(i) If $\mathbb{H}_0$ holds true, then

$$T_n \overset{\mathcal{L}}{\longrightarrow} \mathcal{F}\left(G\left(\cdot\right),\, N\left(0, \sigma_p^2\right)\right)$$

for some functional $\mathcal{F}$, where $G\left(\cdot\right)$ is a tight Gaussian process on $L^\infty[0,1]$ and $\sigma_p^2 < \infty$;
(ii) Under any fixed alternative, $T_n \overset{p}{\longrightarrow} +\infty$.

Theorem 1 shows that T_n converges in distribution to a tight distribution under the null, and diverges to ∞ in probability under any fixed alternative.

Bootstrap implementation

The null asymptotic distribution is unfamiliar. We adopt a bootstrap procedure to obtain critical values for the test. The justification of its asymptotic validity is standard. Specifically, given a bootstrap sample $\left\{c_{it}^*, w_i^*, d_i^{FF*}, d_i^{CF*}\right\}_{i=1}^{n}$ (generated by resampling the original period-t sample $\left\{c_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1}^{n}$), we construct the bootstrap statistic

$$T_n^* = \int_0^1 n_f^* \cdot V^*\left(\tau\right)^2 d\tau,$$

with a re-centered term

$$
\begin{aligned}
V^*(\tau) \quad \equiv \quad & \left\{\left[\widehat{C}_t^*(\tau;\varpi_1)-\widehat{C}_t^*(\tau\hat{p}^*;\varpi_2)\right]-\left[\widehat{C}_t^*(\nu;\varpi_1)-\widehat{C}_t^*(\nu\hat{p}^*;\varpi_2)\right]\right\} \\
& -\left\{\left[\widehat{C}_t(\tau;\varpi_1)-\widehat{C}_t(\tau\hat{p};\varpi_2)\right]-\left[\widehat{C}_t(\nu;\varpi_1)-\widehat{C}_t(\nu\hat{p};\varpi_2)\right]\right\},
\end{aligned}
$$

where: (i) n_f^* is the sample size of $\mathcal{FP}_t^*$, and $\hat{p}^*=\hat{p}_{t,1}^*/\hat{p}_{t,2}^*$, with

$$
\begin{aligned}
& \mathcal{FP}_1^* \equiv \left\{i: d_i^{FF*}=1\right\} \text{ and } \mathcal{FP}_2^* \equiv \left\{i: d_i^{FF*}+d_i^{CF*}=1\right\}, \\
& \hat{p}_{t,j}^* = \frac{\sum_{i=1}^n \mathbf{1}\left(i\in\mathcal{FP}_t^*\right)\mathbf{1}\left(w_i^*=\varpi_j\right)}{\sum_{i=1}^n \mathbf{1}\left(w_i^*=\varpi_j\right)}, \text{ for } j=1,2;
\end{aligned}
$$

(ii) $\widehat{C}_t^*(\tau,\varpi_1)=\widehat{q}_1^*(\tau)$ and $\widehat{C}_t^*(\tau\hat{p}^*;\varpi_2)=\widehat{q}_1^*(\tau\hat{p}^*)+\widehat{q}_2^*(\tau\hat{p}^*)$, with

$$
\begin{aligned}
\left\{\widehat{q}_1^*(\tau),\widehat{q}_2^*(\tau)\right\} &= \underset{q_1,\,q_2}{\operatorname{argmin}} \sum_{i\in\mathcal{FP}_t^*} \rho_\tau\left(c_{it}^*-q_1-q_2\cdot\mathbf{1}\left(w_i^*=\varpi_2\right)\right), \\
\left\{\widehat{q}_1^*(\tau\hat{p}^*),\widehat{q}_2^*(\tau\hat{p}^*)\right\} &= \underset{q_1,\,q_2}{\operatorname{argmin}} \sum_{i\in\mathcal{FP}_t^*} \rho_{\tau\hat{p}^*}\left(c_{it}^*-q_1-q_2\cdot\mathbf{1}\left(w_i^*=\varpi_2\right)\right).
\end{aligned}
$$

4 Identification

In this section we develop regularity conditions for identification of all model primitives

$$
\mathcal{S}\equiv[F(\cdot),H(\cdot),\psi_1(\cdot),\psi_2(\cdot),\alpha/(1+\lambda),\gamma].
$$

As to be shown, none of these conditions requires the LCC to hold. We note that α and λ are not separately identifiable.[7] Therefore, we treat $\alpha/(1+\lambda)$ as a whole for identification.

Besides the observable variables specified in Section 3.1, i.e., $\left\{C_t,Q_t,W,D^{FF},D^{CF}\right\}_{t=1,2}$, we may observe in addition some other characteristics of the principal or/and the agent, denoted by a vector of covariates Z. Though, we make the following arrangements for elaborating our identification strategies: (i) We suppress Z for simplicity of notations, noting that every step involved in the identification strategies would work just fine when conditioned on Z; (ii) We still prioritize the two-period setting, which is our default setting unless otherwise specified.

The following lemma shows that $\mathcal{S}$ is not point identified without further restrictions.

Lemma 4.1 ***(Observational equivalence)*** *Suppose $\mathcal{S}$ satisfies Assumption 3.2. Then, $\mathcal{S}\equiv[F,H,\psi_1,\psi_2,\alpha/(1+\lambda),\gamma]$ is observationally equivalent to $\widetilde{\mathcal{S}}\equiv[\widetilde{F},\widetilde{H},\widetilde{\psi}_1,\widetilde{\psi}_2,\alpha/(1+\lambda),\gamma]$ with $\widetilde{F}(\cdot)=F(\cdot/\xi)$, $\widetilde{H}(\cdot)=H(\cdot/\xi)$, and $\widetilde{\psi}_j(\cdot)=\psi_j(\cdot/\xi)$ for $j=1,2$, for any given constant $\xi>0$.*

To achieve identification, we normalize the supports for θ and e, i.e., $[\underline{\theta},\bar{\theta}]$ and $[\underline{e},\bar{e}]$, to be known.[8]

[7]To see this, note that the economic model remain essentially the same for difference pairs of $\{\alpha,\lambda\}$ as long as the ratio $\alpha/(1+\lambda)$ is the same.

[8]Similar normalization conditions are imposed in Perrigne and Vuong (2011, 2012) for identification.

4.1 Recovering the optimal effort

First, we identify the distribution of period-t effort E_t under FP by adopting Schennach and Hu (2013)'s method, for which we require two effort-related proxies as follows:

Assumption 4.1 *There exist two effort-related proxy variables X_t and Y_t such that*

$$X_t = E_t + V_{t,1} \quad \textit{and} \quad Y_t = m_t(E_t) + V_{t,2},$$

for some unknown function $m_t(\cdot)$, where E_t, $V_{t,1}$, and $V_{t,2}$ are mutually independent with $\mathbb{E}(V_{t,1}) = \mathbb{E}(V_{t,2}) = 0$.

According to Assumption 4.1, X_t can be regarded as an observed measure of E_t with measurement errors. The relation between Y_t and E_t is more flexible as we do not restrict the functional form of $m_t(\cdot)$. For example, agent's effort-related performance is a potential candidate for Y_t. See Cicala (2015) for a discussion on the plausibility of employing cost-related variables to infer the agent's effort. In our empirical study of public transportation service contracts in France, X_t and Y_t are picked to be the share of drivers among all the employees (which primarily consist of drivers and engineers) and the labor fee, respectively. We discuss the justification of such a proxies choice in the empirical section. Note that Assumption 4.1 is less restrictive then requiring two direct measurements of a latent variable, which is imposed by many existing papers to identify various structural models. For instance, Li (2002) requires $m(\cdot)$ to be an identity function.

Denote by $F_{E_t|W}(\cdot|\varpi_j)$ the conditional CDF of period-t effort (under FP) on $W = \varpi_j$. It follows from Theorem 1 of Schennach and Hu (2013) that $F_{E_t|W}(\cdot|\varpi_j)$ and $m_t(\cdot)$ are nonparametrically identifiable from the conditional distribution of (X_t, Y_t) on $W = \varpi_j$, for $j = 1, 2$, except for some rather specific data generating processes (DGP), which impose little restrictions to our model, as discussed (in the proof of Proposition 4.1) in Appendix A.

Once $F_{E_t|W}(\cdot|\varpi_j)$ is identified, we can recover the unobserved effort corresponding to an observed period-t cost c_t under FP conditioned on $W = \varpi_j$ according to Corollary 3.1: (i) c_t corresponds to some $\theta \in \left[\underline{\theta}, \theta_t^{j*}\right]$ such that $c_t = c(\theta; \varpi_j) \equiv H(\theta - e(\theta; \varpi_j))$; (ii) it necessarily holds that $c_t \in \left[c(\underline{\theta}; \varpi_j), c\left(\theta_t^{j*}; \varpi_j\right)\right] \equiv \left[\underline{c}^j, \bar{c}_t^j\right]$; (iii) the mapping from $c_t = c(\theta; \varpi_j)$ to its corresponding effort $e_t = e(\theta; \varpi_j)$ is strictly increasing and bijective from $\left[\underline{c}^j, \bar{c}_t^j\right]$ to $\left[\underline{e}^j, \bar{e}_t^j\right] \equiv \left[e(\underline{\theta}; \varpi_j), e\left(\theta_t^{j*}; \varpi_j\right)\right]$. Consequently, although unobserved, $e_t = e(\theta; \varpi_j)$ can be recovered as

$$e_t = F^{-1}_{E_t|W}(F_{C_t|W}(c_t|\varpi_j)|\varpi_j), \ \forall \ c_t \in [\underline{c}^j, \bar{c}_t^j], \tag{4.1}$$

where $F_{C_t|W}(\cdot|\varpi_j)$ is the conditional CDF of period-t cost under FP on $W = \varpi_j$. Under CR, the effort level is simply zero. We summarize this identification result in the following proposition:

Proposition 4.1 *Let Assumptions 3.1, 3.2 and 4.1 hold. Conditioned on W, the (optimal) effort level associated with any observed cost c_t is identifiable.*

4.2 Identification of the cost function $H(\cdot)$

To identify $H(\cdot)$, we exploit Equation (3.8), which in turn implies that

$$H^{-1}\left(C_t\left(\tau/p_{t,1};\varpi_1\right)\right)=H^{-1}\left(C_t\left(\tau/p_{t,2};\varpi_2\right)\right)+\Delta\tilde{E}_t(\tau),\ \forall\,\tau\in[0,\min\{p_{t,1},p_{t,2}\}] \tag{4.2}$$

with $p_{t,j}\equiv F\left(\theta_t^{j*}\right)$ as previously defined, and $\Delta\tilde{E}_t(\tau)\equiv E_t\left(\tau/p_{t,2};\varpi_2\right)-E_t\left(\tau/p_{t,1};\varpi_1\right)$. Although Condition T.1 suffices for (3.8) and (4.2) to hold, further requirements on W are needed for identification, which are listed in Assumption 4.2.

Assumption 4.2 *The binary variable W satisfies: (i) W is independent of θ; (ii) The disutility from effort is dependent on W such that $\psi_1'(e)\geq\psi_2'(e)$ for all $e\geq\underline{e}$, with equality holding only at $e=\underline{e}$, i.e., $\psi_1'(\underline{e})=\psi_2'(\underline{e})$.*

Assumption 4.2 (i) repeats Condition T.1. Assumption 4.2 (ii) specifies a further requirement: agents with $W=\varpi_1$ incur more marginal disutility than those with $W=\varpi_2$ does, and the two marginal curve $\{\psi_1'(\cdot),\psi_2'(\cdot)\}$ cross only once, at the lower bound $\underline{e}$. Similar single-crossing or finite-crossing conditions are widely imposed for identification purpose in the literature, see, e.g. Chesher (2003), Chernozhukov and Hansen (2005), Heckman et al. (2010), and Torgovitsky (2015), among others.

Assumptions 3.2 and 4.2, in conjunction with the f.o.c. $H'\left(\underline{\theta}-e\left(\underline{\theta};\varpi_j\right)\right)=\psi_j'\left(e\left(\underline{\theta};\varpi_j\right)\right)$, imply the followings: (i) $e\left(\underline{\theta};\varpi_1\right)=e\left(\underline{\theta};\varpi_2\right)=\underline{e}$, hence $\psi_1'\left(e\left(\underline{\theta};\varpi_1\right)\right)=\psi_2'(e\left(\underline{\theta};\varpi_2\right))$. Similarly, $\underline{c}^j=H\left(\underline{\theta}-e\left(\underline{\theta};\varpi_j\right)\right)\equiv H\left(\underline{\theta}-\underline{e}\right)=\underline{c}$ for $j=1,2$; (ii) $e\left(\theta;\varpi_1\right)<e\left(\theta;\varpi_2\right)$ for any $\theta>\underline{\theta}$; (iii) Regarding θ_t^{j*}, i.e., the cutoff type (between choosing FP and choosing CR for period-t) conditioned $W=\varpi_j$, we have $\theta_t^{1*}\leq\theta_t^{2*}$. Consequently, $p_{t,1}\leq p_{t,2}$, for $t=1,2$.

A key step for identifying $H(\cdot)$ is to establish for the term $\Delta\tilde{E}_t(\tau)$ in (4.2) that

$$\Delta\tilde{E}_t(\tau)>0\text{ for all }\tau\in(0,p_{t,1}]\,, \tag{4.3}$$

which we show in the proof of Proposition 4.2 in Appendix A. Recall that, $F_{C_t|W}(\cdot|\varpi_j)$, the conditional CDF of period-t cost under FP on $W=\varpi_j$, is with support $\left[\underline{c},\bar{c}_t^j\right]$. So, for any $c\in\left[\underline{c},\bar{c}_t^1\right]$, there exists a unique quantile level $\tau_0(c)\in[0,1]$ such that $c=C_t\left(\tau_0(c);\varpi_1\right)$.[9] If $c=\underline{c}$, it holds, according to Corollary 3.1 (i), that $\tau_0(c)=0$ and that $H^{-1}(\underline{c})$ is immediately identified as $H^{-1}(\underline{c})=\underline{\theta}-\underline{e}$. If $c\in\left(\underline{c},\bar{c}_t^1\right]$, it holds, again according to Corollary 3.1 (i), that $\tau_0(c)\in(0,1]$. Moreover, for $c\in\left(\underline{c},\bar{c}_t^1\right]$, it follows from (4.2) and (4.3) that $C_t\left(\tau_0(c);\varpi_1\right)=C_t\left(\tau_0(c)\cdot p_{t,1}/p_{t,2};\varpi_2\right)+\Delta\tilde{E}_t\left(\tau_0(c)\cdot p_{t,1}\right)>C_t\left(\tau_0(c)\cdot p_{t,1}/p_{t,2};\varpi_2\right)>\underline{c}$. It follows from the same logic that there exists $\tau_1(c)\in(0,\tau_0(c))$ such that $C_t\left(\tau_0(c)\cdot p_{t,1}/p_{t,2};\varpi_2\right)=C_t\left(\tau_1(c);\varpi_1\right)>C_t\left(\tau_1(c)\cdot p_{t,1}/p_{t,2};\varpi_2\right)$. Recursively, for any $c\in\left(\underline{c},\bar{c}_t^1\right]$, we can establish a strictly decreasing sequence of quantiles $\{\tau_k(c)\}_{k=0}^{\infty}$ such that $C_t\left(\tau_k(c).p_{t,1}/p_{t,2};\varpi_2\right)=C_t\left(\tau_{k+1}(c);\varpi_1\right)>$

[9] We implicitly assume $F_{C_t|W}(\cdot|\varpi_j)$ to be continuous and strictly increasing on its support.

$C_t\left(\tau_{k+1}(c).p_{t,1}/p_{t,2};\varpi_2\right)$, with a initial condition that $C_t\left(\tau_0(c);\varpi_1\right)=c$. Iterating (4.2) $m+1$ times according to $\{\tau_k(c)\}_{k=0}^{\infty}$ yields:

$$\sum\nolimits_{k=0}^{m}\Delta\tilde{E}_t(\tau_k(c).p_{t,1})=H^{-1}(c)-H^{-1}\left(C_t\left(\tau_m(c).p_{t,1}/p_{t,2};\varpi_2\right)\right).$$

In the proof of Proposition 4.2 (in Appendix A), we show the followings: (i) The decreasing sequence $\{\tau_k\left(c\right)\}_{k=0}^{\infty}$ converges to 0; (ii) $\sum_{k=0}^{\infty}\Delta\tilde{E}_t(\tau_k(t).p_{t,1})<\infty$. Consequently, taking limit of $m\to\infty$ over the equation above and rearranging terms yields

$$H^{-1}\left(c\right)=H^{-1}\left(C_t\left(0;\varpi_2\right)\right)+\sum\nolimits_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c).p_{t,1}\right)=\underline{\theta}-\underline{e}+\sum\nolimits_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c).p_{t,1}\right)\quad(4.4)$$

for any $c\in\left(\underline{c},\bar{c}_t^1\right]$. Note that the whole sequence $\{\tau_k(c)\}_{k=0}^{\infty}$ is identifiable, because $F_{C_t|W}\left(\cdot|\varpi_j\right)$ are directly identifiable as both C_t and W are observed. And $\Delta\tilde{E}_t\left(\tau_k\cdot p_{t,1}\right)=E_t\left(\tau_k.p_{t,1}/p_{t,2};\varpi_2\right)-E_t\left(\tau_k;\varpi_1\right)$ is identifiable by Proposition 4.1. Therefore, (4.4) identifies $H^{-1}\left(c\right)$ for any $c\in\left(\underline{c},\bar{c}_t^1\right]$. Note that Guerre et al. (2009) and D'Haultfoeuille and Février (2019) achieve identification by exploit quantile relations that are similar to (4.4). We formalize the identification result for $H\left(\cdot\right)$ in the following proposition:

Proposition 4.2 *Let Assumptions 3.1, 3.2, 4.1 and 4.2 hold.* $H^{-1}(\cdot)$ *is nonparametrically identified on* $[\underline{c},\bar{c}_t^1]$ *as*

$$H^{-1}(c)=\begin{cases}\underline{\theta}-\underline{e}, & \textit{for } c=\underline{c},\\ \underline{\theta}-\underline{e}+\sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right), & \textit{for } c\in\left(\underline{c},\bar{c}_t^1\right].\end{cases}\quad(4.5)$$

As shown in the proof of Proposition 4.2, the single-crossing condition of Assumption 4.2 (ii) plays a crucial role: It helps to locate the limiting pointing of the sequence $\{\tau_k\left(c\right)\}_{k=0}^{\infty}$, which eventually leads to the identification result above. Nevertheless, it is worth noting that whether the crossing happens at a boundary point ($\underline{e}$ as in Assumption 4.2 (ii)) or not is nonessential, in the sense that $H\left(\cdot\right)$ can be identified (on certain range) by similar iterative steps when Assumption 4.2 (ii)'s single-crossing at the boundary is altered to single-crossing at an interior point. We formalize this alternative identification result in Appendix D.1.

4.3 Identification of other model primitives

Based on identification of $H^{-1}\left(\cdot\right)$ on $\left[\underline{c},\bar{c}_t^1\right]$, we proceed to identify other model primitives, in the following order: the type distribution, the disutility functions, $\alpha/\left(1+\lambda\right)$, and γ.

The type distribution

Conditioned on $W=\varpi_1$, a period-t cost $c\in\left[\underline{c},\bar{c}^1\right]$ necessarily corresponds to $\theta\in\left[\underline{\theta},\theta_t^{1*}\right]$ and FP being chosen for that period, while $c>\bar{c}_t^1$ corresponds to $\theta\in\left(\theta_t^{1*},\bar{\theta}\,\right]$ and CR. Specifically, conditioned on $W = \varpi_1$, for any realized period-t cost c, it holds for the corresponding θ that

$$\theta = \begin{cases} H^{-1}(c) + F^{-1}_{E_t|W}\left(F_{C_t|W}\left(c|\varpi_1\right)|\varpi_1\right), \text{ for } c \in \left[\underline{c}, \bar{c}^1_t\right]; \\ H^{-1}(c), \text{ for } c \in (\bar{c}^1_t, \bar{c}]. \end{cases} \tag{4.6}$$

Given the identification results in Sections 4.1 and 4.2, we can recover the corresponding θ for any realized period-t cost $c \in \left[\underline{c}, \bar{c}^1_t\right]$ by (4.6). Consequently, we identify the truncated distribution of θ on $\left[\underline{\theta}, \theta^{1*}_t\right]$ (i.e., conditional on $\theta \in \left[\underline{\theta}, \theta^{1*}\right]$), for which we denote by $G(\cdot)$ and $g(\cdot)$ its CDF and pdf. We can then identify the (unconditional) CDF and pdf of θ on $\left[\underline{\theta}, \theta^{1*}_t\right]$ as

$$F(\cdot) = G(\cdot)F\left(\theta^{1*}_t\right) \quad \text{and} \quad f(\cdot) = g(\cdot)F\left(\theta^{1*}_t\right), \text{ respectively,} \tag{4.7}$$

with $F\left(\theta^{1*}_t\right) = \mathbb{E}\left(D^{FF}|W = \varpi_1\right)$ for $t = 1$, or $\mathbb{E}\left(D^{CF} + D^{FF}|W = \varpi_1\right)$ for $t = 2$, directly identifiable. However, without additional assumptions, it is impossible to identify $F(\cdot)$ or $f(\cdot)$ on $\left(\theta^{1*}_t, \bar{\theta}\right]$, since $H^{-1}(\cdot)$ is not identified on $\left(\bar{c}^1_t, \bar{c}\right]$.

The disutility functions

It follows from (4.1) that, conditioned on $W = \varpi_j$, the cost c corresponding to any effort $e \in \left[\underline{e}, \bar{e}^j_t\right]$ in period-t is given by $c = F^{-1}_{C_t|W}\left(F_{E_t|W}\left(e|\varpi_j\right)|\varpi_j\right) \in \left[\underline{c}, \bar{c}^j_t\right]$, which, together with the f.o.c. (3.3), imply that, for any $e \in \left[\underline{e}, \bar{e}^j_t\right]$,

$$\psi'_j(e) = H'\left(H^{-1}\left(F^{-1}_{C_t|W}\left(F_{E_t|W}\left(e|\varpi_j\right)|\varpi_j\right)\right)\right) = \frac{1}{H^{-1\prime}\left(F^{-1}_{C_t|W}\left(F_{E_t|W}\left(e|\varpi_j\right)|\varpi_j\right)\right)} \tag{4.8}$$

Given identification results already established, $\psi_j(\cdot)$ is identified by its differential equation (4.8), together with the location normalization condition $\psi_j(0) = 0$ by Assumption 3.2 (ii).

The ratio $\alpha/(1+\lambda)$ and intertemporal preference γ

The identification of $\alpha/(1+\lambda)$ is based on (3.6). Rearranging terms for (3.6) yields

$$\begin{aligned} \frac{\alpha}{1+\lambda} &= 1 - \frac{H\left(\theta^{1*}_2\right) - \bar{b}^1}{H'\left(\theta^{1*}_2 - e\left(\theta^{1*}_2; \varpi_1\right)\right)} \frac{f\left(\theta^{1*}_2\right)}{F\left(\theta^{1*}_2\right) - F\left(\theta^{1*}_1\right)} \\ &= 1 - \frac{\underline{c}^1_{CC} - \bar{b}^1}{H'\left(H^{-1}\left(\bar{c}^1_2\right)\right)} \frac{f(\theta^{1*}_2)}{p_{2,1} - p_{1,1}} \end{aligned} \tag{4.9}$$

with $\underline{c}^1_{CC} \equiv H\left(\theta^{1*}_2\right)\left(> H\left(\theta^{1*}_2 - \bar{e}^1_2\right) = \bar{c}^1_2\right)$ being the lower bound for period-2 cost under CC, conditioned on $W = \varpi_1$. Since all terms of (4.9) are either directly identifiable from the observed variables or already shown identifiable, (4.9) identifies $\alpha/(1+\lambda)$.

The identification of γ is based on (3.5), as follows:

$$\gamma = \frac{\bar{b}^1 - \underline{b}^1}{\bar{b}^1 - H\left(\theta^{1*}_1 - \bar{e}^1_1\right) - \psi_1\left(\bar{e}^1_1\right)} = \frac{\bar{b}^1 - \underline{b}^1}{\bar{b}^1 - \bar{c}^1_1 - \psi_1\left(\bar{e}^1_1\right)} \tag{4.10}$$

where $\bar{e}_1^1 \in \left[\underline{e}, \bar{e}_2^1\right]$ is identified as the solution to the f.o.c. $\psi_1'\left(\bar{e}_1^1\right) = H'\left(H^{-1}\left(\bar{c}_1^1\right)\right)$. Alternatively, according to (4.1) and the fact that $F_{C_1|W}\left(\bar{c}_1^1|\varpi_1\right) = 1$, $\bar{e}_1^1$ is identified as $\bar{e}_1^1 = \inf\left\{e : F_{E_1|W}\left(e|\varpi_1\right) = 1\right\}$.

4.4 Summary of identification results and further discussion

As shown in Sections 4.1-4.3, the identification strategies for $H(\cdot)$, $F(\cdot)$ and $(\psi_1(\cdot), \psi_2(\cdot))'$ can be based on observations solely from either period-1 or period-2, and they work the same way regardless of which period is used. However, the identifiable intervals (i.e., sub-domains) do depend on which period is used: Based on period-t observations, $H^{-1}(\cdot)$ is nonparametrically identified on $\left[\underline{c}, \bar{c}_t^1\right]$. Consequently, $H(\cdot)$, $F(\cdot)$ and $(\psi_1(\cdot), \psi_2(\cdot))'$ are nonparametrically identified on $\left[\underline{\theta} - \underline{e},\ \theta_t^{1*} - \bar{e}_2^1\right]$, $\left[\underline{\theta}, \theta_t^{1*}\right]$ and $\left[\underline{e}, \bar{e}_t^1\right]$, respectively.

Recall the following facts: (i) $\bar{c}_t^1 = c\left(\theta_t^{1*}; \varpi_1\right)$ by definition; $c(\theta; \varpi_1) = H(\theta - e(\theta; \varpi_1))$ is strictly increasing in θ (according to Corollary 3.1 (i)); (iii) $\theta_2^{1*} > \theta_1^{1*}$ (according to Corollary 3.1 (ii)). These facts together immediately imply that $\bar{c}_2^1 > \bar{c}_1^1$. Therefore, period-2 observations identify $H^{-1}(\cdot)$ on a larger interval than period-1 observations do (i.e., $[\underline{c}, \bar{c}_2^1] \supset [\underline{c}, \bar{c}_1^1]$). Consequently, period-2 observations also identify $H(\cdot)$, $F(\cdot)$ and $(\psi_1(\cdot), \psi_2(\cdot))'$ on larger intervals. Note that the identification of $\alpha/(1+\lambda)$ and γ require observations from both periods. We summarize identification results established so far in the following theorem:

Theorem 2 *Let Assumptions 3.1, 3.2, 4.1 and 4.2 hold. For the two-period setting, $H^{-1}(\cdot)$ is nonparametrically identified on $\left[\underline{c}, \bar{c}_2^1\right]$. Consequently, $H(\cdot)$, $F(\cdot)$ and $(\psi_1(\cdot), \psi_2(\cdot))'$ are nonparametrically identified on $\left[\underline{\theta} - \underline{e},\ \theta_2^{1*} - \bar{e}_2^1\right]$, $\left[\underline{\theta}, \theta_2^{1*}\right]$ and $\left[\underline{e}, \bar{e}_2^1\right]$, respectively. In addition, both $\alpha/(1+\lambda)$ and γ are identified.*

Without imposing further restrictions, it is impossible to identify $H^{-1}(\cdot)$ on $\left(\bar{c}_2^1, \bar{c}\right]$ or equivalently $H(\cdot)$ on $\left(\theta_2^{1*} - \bar{e}_2^1, \overline{\theta}\right]$, or to identify $F(\cdot)$ on $\left(\theta_2^{1*}, \overline{\theta}\right]$. This is because $\left(\theta_2^{1*}, \overline{\theta}\right]$ is the region for θ where the ϖ_1-class agent would choose CR rather than FP, $\left(\bar{c}_2^1, \bar{c}\right]$ and $\left(\theta_2^{1*} - \bar{e}_2^1, \overline{\theta}\right]$ being the corresponding regions for the period-2 cost and $\theta - e$, respectively. And there is a lack of (structural) information under CR regions. After all, under CR, we only observe the cost, and the effort is constantly zero (rather than being strict monotonic under FP). And the identification strategy for $H^{-1}(\cdot)$, essentialized into Equations (4.2) - (4.5), only works on the region where both ϖ_1- and ϖ_2-classes choose FP, thus is not applicable to the region $\left(\theta_2^{1*} - \bar{e}_2^1, \overline{\theta}\right]$.

In what follows, we discuss parametric identification of $H^{-1}(\cdot)$ on its entire domain $[\underline{c}, \bar{c}]$. Under a parametric specification $H^{-1}(\cdot) = H^{-1}(\cdot; \beta)$ for some finite dimensional $\beta \in \mathbb{R}^{k_\beta}$, note the followings: (i) The parameterization guarantees an unique extrapolation of $H^{-1}(\cdot)$ from the information-rich region $\left[\underline{c}, \bar{c}^1\right]$ to $\left(\bar{c}^1, \bar{c}\right]$; (ii) (4.4) still holds for all $c \in \left[\underline{\theta} - \underline{e}, \theta_2^{1*} - \bar{e}_2^1\right]$, and suffices to identify β. Alternatively, β can be identified by

$$H^{-1}\left(C_t\left(a_l; \varpi_1\right); \beta\right) = H^{-1}\left(C_t\left(a_l \cdot p_{t,1}/p_{t,2}; \varpi_2\right); \beta\right) + \Delta\tilde{E}_t\left(a_l \cdot p_{t,1}\right) \tag{4.11}$$

for a pre-selected grids $\{a_1, a_2, ..., a_L\} \subset (0, 1)$ with L large enough, together with the condition $H^{-1}(\underline{c}; \beta) = \underline{\theta} - \underline{e}$. Typically, $L \geq k_\beta$ is required for, but dose not necessarily guarantee, identification of β. Detailed conditions for identification of nonlinear parametric models (i.e., uniqueness of solution to a system of nonlinear equations) is case dependent; (iii) Once $H(\cdot)$ is identified on $[\underline{c}, \bar{c}]$, $F(\cdot)$ is identified on $\left[\underline{\theta}, \overline{\theta}\right]$ according to (4.6); (iv) Parameterization of various cost functions is a widely adopted in structural analysis in the literature. For example, Luo et al. (2018a) parameterize a cost function to identify the truncated distribution of consumer type.

It is possible, for the most part, to straightforwardly extend the econometric analyses to T-period settings with $T > 2$. Similar to the two-period setting, for $T > 2$, the equilibrium is roughly characterized as follows: Conditioned on W, there would be T cutoff types that divide agents into $T+1$ categories from the most efficient to the least efficient. The most efficient agents choose FP for all periods, the second-most efficient ones choose FP for all periods except for period-1, the third-most efficient ones choose FP for all periods except for period-1 and 2, and the least efficient agents choose CR for all periods. Based on these equilibrium characteristics, $H^{-1}(\cdot)$, $F(\cdot)$, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ can be identified by following essentially the same steps as those proposed in Section 4 and utilizing period-T (i.e. the last period) observations. However, the identification of $\alpha/(1+\lambda)$ and γ requires deriving equations parallel (3.5) and (3.6) (of Corollary 3.1), which is nontrivial for $T > 2$.

5 Estimation

In this section, we propose estimation procedures for the model primitives in the two-period setting, regarding which we note the followings: (i) The proposed procedures are fully nonparametric. Nevertheless, they can be flexibly adapted to parametric or semiparametric specifications if needed/preferred (perhaps when the sample size is not that large); (ii) The proposed procedures are readily extendable to the one-period setting; (iii) As previously mentioned, we may observe other covariates Z, on which some of the model primitives are potentially dependent. With Z being suppressed, the proposed procedures are implicitly conditioned on $Z = z$ for a given z, and provide a basis for developing more sophisticated estimation procedures that incorporate Z, nonparametrically or parametrically.[10] The estimation procedures are mainly based on the period-2 sample $\left\{c_{i2}, q_{i2}, x_{i2}, y_{i2}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1}^n$, and closely follow the identification strategies in Section 4.1 - 4.3, as follows.

[10]Depending on the sample size, it may be preferable to incorporate Z parametrically to avoid "curse of dimensionality".

Period-2 effort distribution

Denote by $f_{E_2|W}(\cdot|\varpi_j)$, $f_{V_{2,1}|W}(\cdot|\varpi_j)$ and $f_{V_{2,2}|W}(\cdot|\varpi_j)$ the conditional pdf's of E_2, $V_{2,1}$ and $V_{2,2}$ on $W = \varpi_j$, respectively. We estimate these pdf's jointly with $m_2(\cdot)$ by the sieve maximum likelihood estimation (SMLE) proposed in Shen (1997). Specifically, for $j = 1, 2$ each, the SMLE estimator $\left(\widehat{f}_{E_2|W}(\cdot|\varpi_j), \widehat{f}_{V_{2,1}|W}(\cdot|\varpi_j), \widehat{f}_{V_{2,2}|W}(\cdot|\varpi_j), \widehat{m}_2(\cdot)\right)$ maximizes the following log-likelihood objective function

$$\sum_{i\in\mathcal{FP}_2^j} \ln \int f_{V_{2,1}|W}(x_{i2} - e|\varpi_j)\, f_{V_{2,2}|W}(y_{i2} - m_2(e)|\varpi_j)\, f_{E_2|W}(e|\varpi_j)\, de,$$

subject to the following constraints:

(i) $F_{E_2|W}(\cdot|\varpi_j)$, $f_{V_{2,1}|W}(\cdot|\varpi_j)$ and $f_{V_{2,2}|W}(\cdot|\varpi_j)$ are all nonnegative, and integrate to 1;

(ii) $\int v f_{V_{2,1}|W}(v|\varpi_j)\, dv = \int v f_{V_{2,2}|W}(v|\varpi_j)\, dv = 0$ (to satisfy Assumption 4.1).

For easy implementation of constraints (i) and (ii) above, we recommend to approximate the density functions $f_{V_1|W}$, $f_{V_2|W}$ and $f_{E_0|W}$ by nonlinear sieves

$$f_{V_{2,1}|W}(\cdot|\varpi_j) = \left[\sum_{k=0}^{k_{1n}} a_{jk} q_k(\cdot)\right]^2,\ f_{V_{2,2}|W}(\cdot|\varpi_j) = \left[\sum_{k=0}^{k_{2n}} b_{jk} q_k(\cdot)\right]^2,\ f_{E_2|W}(\cdot|\varpi_j) = \left[\sum_{k=0}^{k_{3n}} c_{jk} q_k(\cdot)\right]^2,$$

based on *orthogonal Hermite* basis functions $q_k(\iota) = \sqrt{\frac{1}{\sqrt{\pi k}\,!2^k}} H_k(\iota) e^{-\frac{\iota^2}{2}}$, with $H_0(\iota) = 1$, $H_1(\iota) = 2\iota$ and $H_{k+1}(\iota) = 2\iota H_k(\iota) - 2kH_{k-1}(\iota)$, where k_{1n}, k_{2n} and k_{3n} are the chosen numbers of basis functions. For comprehensive studies on the implementation and properties of sieve estimations, see Shen (1997), Ai and Chen (2003), Chen (2007), Carroll et al. (2010), and Chen et al. (2014), among others.

The (inverse) cost function

Motivated by (4.2), we construct the square-distance based operator

$$Q_n\left(H^{-1}\right) = \int_0^1 \left[H^{-1}\left(\widehat{C}_2(\tau\cdot\hat{p}_{2,1}/\hat{p}_{2,2};\varpi_2)\right) - H^{-1}\left(\widehat{C}_2(\tau;\varpi_1)\right) + \widehat{\Delta\widehat{E}_2}(\tau\cdot\hat{p}_{2,1})\right]^2 d\tau,$$

where, for any given τ, $\widehat{C}_2(\tau;\varpi_j)$ and $\widehat{E}_2(\tau;\varpi_j)$ are consistent estimators for $C_2(\tau;\varpi_j)$ and $E_2(\tau;\varpi_j)$, respectively, and $\widehat{\Delta\widehat{E}_2}(\tau\cdot\hat{p}_{2,1}) = \widehat{E}_2(\tau\cdot\hat{p}_{2,1}/\hat{p}_{2,2};\varpi_2) - \widehat{E}_2(\tau;\varpi_1)$. $\hat{p}_{2,j}$, for $j = 1, 2$, are as specified in (3.11), i.e., $\hat{p}_{2,j} = \sum_{i=1}^n \mathbf{1}(i\in\mathcal{FP}_2)\,\mathbf{1}(w_i = \varpi_j)/\sum_{i=1}^n \mathbf{1}(w_i = \varpi_j)$.

We propose the following sieve estimator for $H^{-1}(\cdot)$:

$$\widehat{H}^{-1}(\cdot) = \underset{H^{-1}\in\mathcal{H}_n:\ H^{-1}(\underline{c}) = \underline{\theta} - \underline{e}}{\operatorname{argmin}} Q_n\left(H^{-1}\right), \tag{5.1}$$

where $\mathcal{H}_n$ is a sieve approximating space with a finite dimension k_n, such that $k_n \to \infty$ as $n \to \infty$. Regarding the implementation of the sieve estimation (5.1) above, we note the followings: (i) We recommend to use linear sieve, where $\mathcal{H}_n$ is a linear space spanned by k_n basis functions, for example, B-splines as a population choice; (ii) For a given τ, $\widehat{C}_2(\tau; \varpi_j)$ can be constructed as the τ'th percentile of $\{c_{i,2}\}_{i \in \mathcal{FP}_2^j}$, with $\mathcal{FP}_2^j$ specified by (3.2); (iii) $\widehat{E}_2(\tau; \varpi_j)$ is calculated based on $\widehat{f}_{E_2|W}(\cdot|\varpi_j)$. A possible way to process is to simulate a sample according to $\widehat{f}_{E_2|W}(\cdot|\varpi_j)$ (as the pdf), and then to construct $\widehat{E}_2(\tau; \varpi_j)$ as the τ'th percentile of the simulated sample.

Type distribution and the disutility functions

For each c_{i2}, $i = 1, ..., n$, we can now recover (i.e., estimate) the associated e_{i2} and θ_i, so that we effectively obtain $\{c_{i2}, e_{i2}, \theta_i\}_{i=1}^n$ afterwards. To recover effort e_{i2}, we have

$$e_{i2} = \begin{cases} 0, \text{ if } i \notin \mathcal{FP}_2 \text{ (i.e., } i \notin \mathcal{FP}_2^1 \text{ and } i \notin \mathcal{FP}_2^2); \\ \widehat{F}_{C_2|W}(c_{i2|\varpi_j})\text{'th quantile of } \widehat{f}_{E_2|W}(\cdot|\varpi_j), \text{ if } i \in \mathcal{FP}_2^j \text{ for } j = 1 \text{ or } 2. \end{cases}$$

To recover θ_i, we have $\theta_i = \widehat{H}^{-1}(c_i) + e_{i2}$. Based on $\{\theta_i\}_{i=1}^n$, $F(\cdot)$ and $f(\cdot)$ can be estimated as the empirical CDF and kernel density estimator, respectively.

To estimate $\psi_j(\cdot)$, for $j = 1, 2$, note that, according to the f.o.c. (3.3), we have

$$\psi_j'(e_{i2}) = H'(c_{i2}) = \frac{1}{H^{-1\prime}(c_{i2})}, \text{ for all } i \in \mathcal{FP}_2^j.$$

Consequently, based on $\{c_{i2}, e_{i2}, \theta_i\}_{i \in \mathcal{FP}_2^j}$, we propose to estimate $\psi_j(\cdot)$ by the following sieve minimum distance estimator:

$$\widehat{\psi}_j(\cdot) = \underset{\psi_j \in \Psi_n:\ \psi_j(0)=0}{\operatorname{argmin}} \sum_{i \in \mathcal{FP}_2^j} \left[\psi_j'(e_{i2}) - \frac{1}{\widehat{H}^{-1\prime}(c_{i2})} \right]^2,$$

where Ψ_n is a sieve approximating space with a finite dimension q_n, such that $q_n \to \infty$ as $n \to \infty$. Like for $\mathcal{H}_n$, we recommend to use linear sieve for Ψ_n, such as B-splines, which facilitate implementing the involved derivatives.

Parameters $\alpha/(1+\lambda)$ and γ

Following (4.9), $\alpha/(1+\lambda)$ is estimated by

$$\widehat{\alpha/(1+\lambda)} = 1 - \frac{\underline{c}_{CC}^1 - \bar{b}^1}{H'\left(H^{-1}\left(\bar{c}_2^1; \hat{\beta}\right); \hat{\beta}\right)} \frac{\widehat{f}\left(\hat{\theta}_2^{1*}\right)}{\hat{p}_{2,1} - \hat{p}_{1,1}},$$

where $\hat{p}_{t,1} = \sum_{i=1}^n \mathbf{1}(i \in \mathcal{FP}_t) \mathbf{1}(w_i = \varpi_1) / \sum_{i=1}^n \mathbf{1}(w_i = \varpi_1)$ and $\hat{\theta}_2^{1*} = \widehat{F}^{-1}(\hat{p}_{2,1})$.

Following (4.10), the intertemporal preference γ is estimated by

$$\hat{\gamma} \;=\; \frac{\bar{b}^1 - \underline{b}^1}{\bar{b}^1 - \bar{c}_1^1 - \widehat{\psi}_1\left(\hat{e}_1^1\right)},$$

where $\hat{e}_1^1 = \widehat{F}_{E_1|W}^{-1}\left(1-\varepsilon_n|\varpi_1\right)$ is an estimator of $\bar{e}_1^1$, with tuning parameter $\varepsilon_n > 0$ such that $\varepsilon_n \downarrow 0$ as $n \to \infty$, based on the fact that $\bar{e}_1^1 = \min\left\{e : F_{E_1|W}\left(e|\varpi_1\right) = 1\right\}$.

Proposition 5.1 establishes regularity conditions for consistency of the main estimators.

Proposition 5.1 *Let Assumptions Assumptions 3.1, 3.2, 4.1 and 4.2 hold. In addition, let the following conditions hold: (i) $H^{-1}(\cdot)$ is identified on $[\underline{c}, \bar{c}]$, and lies in a bounded Sobolev subspace $\mathcal{H}$. (ii) As $n \to \infty$, $k_n \to \infty$ so that $\mathcal{H}_n$ becomes dense in $\mathcal{H}$ at the limit; (iii) $\left\{\hat{p}_{2,1}, \hat{p}_{2,2}, \widehat{C_2}\left(\tau;\varpi_1\right),\ \widehat{C_2}\left(\tau \cdot p_{2,1}/p_{2,2};\varpi_2\right), \widehat{E}\left(\tau;\varpi_1\right),\ \widehat{E}_2\left(\tau \cdot p_{2,1}/p_{2,2};\varpi_2\right)\right\}$ for $\tau \in [0,1]$ are consistent. Then $\widehat{H}^{-1}(\cdot)$, $\widehat{F}(\cdot)$, $\widehat{\psi}_1(\cdot)$ and $\widehat{\psi}_2(\cdot)$ are consistent under the L^2-norm, and $\widehat{\alpha/(1+\lambda)}$ and $\hat{\gamma}$ are consistent.*

6 Monte Carlo Simulations

In this section, we study the finite sample performance of the proposed test for LCC (in Section 3) and estimation procedure (in Section 5), focusing on the two-period setting like previous parts of the paper.

6.1 Data generating process

We consider a series of data generating processes (DGPs) designed as follows: θ is set to follow a uniform distribution on $[1,2]$, $W = \varpi_j$ for $j = 1,2$ with $\Pr(W = \varpi_1) = 0.4$. The cost function is specified as $H(\theta - e) = \beta_1(\theta - e) + \beta_2(\theta - e)^2$. The disutility conditional on $W = \varpi_j$ is specified as $\psi_j\left(e\right) = \kappa_1^j e + \kappa_2^j e^2$ for $j = 1,2$. We set $\beta_1 = 1$, and $\beta_2 = 0$, 1, 1.5, or 2, depending on the specific simulations. Given a specific value-combination of (β_1, β_2), we set the parameters of the disutility functions $(\kappa_1^1, \kappa_2^1, \kappa_1^2, \kappa_2^2)$ accordingly, such that all identification conditions, in particular Assumption 4.2 (ii), are satisfied.[11] Last, we set $\lambda = 0.3$, $\gamma = 0.5$ and $\alpha = 0.5$.

To generate a random sample from a given DGP (i.e., a given $(\beta_1, \beta_2, \kappa_1^1, \kappa_2^1, \kappa_1^2, \kappa_2^2)$ setting), we execute the following steps: (i) For $j = 1,2$, we numerically solve for the equilibrium outcomes $\{\theta_1^{j*}, \theta_2^{j*}, \bar{b}^j, \underline{b}^j\}$ as the solution to a system of nonlinear equations, consisting of (3.4) - (3.6) (from Corollary 3.1) and (A.23) (from Appendix A);[12] (ii) To generate contract/observation i,

[11] For $\beta_2 = 1$, 1.5 and 2, we set $(\kappa_1^1, \kappa_2^1, \kappa_1^2, \kappa_2^2)$ to be $(2,1,1,3)$, $(2,2,0.86,4)$, and $(2,2,0.5,4)$, respectively. When $\beta_2 = 0$, which corresponds to the case of LCC being true (for studying the size performance of the LCC test), the parameters of $\psi_j\left(\cdot\right)$'s need to be specified a bit differently, as we discuss in Section 6.2.

[12] Although (3.4) - (3.6), together with all regularity conditions previously specified, suffice to identify all model primitives (as shown in Section 4), reversely solving for the equilibrium outcomes from a given fully specified model requires (A.23) in addition.

Table 1: LCC test: simulated rejection probabilities

	Size	Power		
Sample size	$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 1.5$	$\beta_2 = 2$
$n = 250$	0.002	0.162	0.452	0.800
$n = 500$	0.024	0.424	0.834	0.996
$n = 1000$	0.056	0.854	0.998	1.000

Notes: The significance level is targeted at 5%.

we randomly draw θ_i and w_i according to corresponding distributions specified in the DGP; (iii) For $w_i = \varpi_j$, we compare θ_i with the cutoffs $\{\theta_1^{j*}, \theta_2^{j*}\}$ to decide whether contract i is FF, CF, or CC, and then assign the values for d_i^{FF} and d_i^{CF} accordingly; (iv) For $w_i = \varpi_j$, we set e_{it} to be the solution to the f.o.c. $H'(\theta_i - e) = \psi_j'(e)$ if FP is chosen for period-t, or we set $e_{it} = 0$ if CR is chosen for period-t, for $t = 1, 2$; (v) We set $c_{it} = H(\theta_i - e_{it})$; (vi) Regarding subsidy q_{it}, for $w_i = \varpi_j$, we set $q_{it} = c_{it}$ for both $t = 1, 2$ if contract i is CC, or we set $q_{i,1} = c_{i,1}$ and $q_{i,2} = \bar{b}^j$ if CF, or we set $q_{i,1} = q_{i,2} = \underline{b}^j$ if FF; (vii) In addition, for $w_i = \varpi_j$, we generate the two proxies x_{it} and y_{it} as $x_{it} = e_{it} + v_{t,1}$ and $y_{it} = \zeta_1^j e_{it} + \zeta_2^j (e_{it})^2 + v_{t,2}$, where $v_{t,1}$ and $v_{t,2}$ are independently drawn from the standard normal distribution, and $\zeta_1^1 = \zeta_2^1 = 1$, and $\zeta_1^2 = \zeta_2^2 = 2$, for $t = 1, 2$; (viii) We independently repeat steps (ii) - (vii) n times to collect a sample of size n $\left\{c_{it}, q_{it}, x_{it}, y_{it}, w_i, d_i^{FF}, d_i^{CF}\right\}_{i=1,\ldots,n,\ t=1,2}$.

For all Monte Carlo experiments, we consider sample sizes $n = 250$, 500 and 1000. All simulation results are based on 500 replications. For implementing the LCC test, the critical values are based on 500 bootstrap evaluations.

6.2 The LCC test: size and power performance

To study the size performance of the LCC test, we set $\beta_2 = 0$ so that the null hypothesis of LCC holds true. As shown in Appendix B.1, under the LCC, the coefficients of the disutility functions κ_1^j and κ_2^j cannot be separately identified. So we normalize the disutility function parameters such that $\psi_1(e) = e^2$, and $\psi_2(e) = \kappa e^2$ with $\kappa = 0.7$. To study the power performance, we set $\beta_2 = 1$, 1.5, and 2, and set all other parameters strictly according to Section 6.1's design.

Table 1 reports the rejection frequencies with a nominal size of 5%. From Table 1 we have the following observations: (i) A sample size of 250 seems to be somewhat too small for the current design, as indicated by the simulated sizes being noticeably below the targeted size of 0.05 at $n = 250$; (ii) At $n = 500$, the test provides reasonably good size control. At $n = 1000$, the simulated size becomes very close to the targeted size; (iii) $\mathbb{H}_0$ is true when $\beta_2 = 0$, but is false when $\beta_2 = 1$, 1.5, and 2. The rejection probabilities increase noticeably as β_2 deviates away from 0 at all sample sizes, and they increase as the sample size increases. This indicates good power performance.

Table 2: Estimation results of model parameters

	$\gamma = 0.5$			$\alpha = 0.5$		
Sample size	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
Mean	0.302	0.437	0.472	0.713	0.587	0.534
Std.	0.002	0.002	0.001	0.442	0.364	0.226
RMSE	0.198	0.063	0.028	0.491	0.374	0.228

Notes: For each of the sample size, the number of replications is 500.

6.3 Estimation performance

We focus on the DGP with $\beta_2 = 1$ for studying the performance of the nonparametric estimation procedures proposed in Section 5. For the sieve spaces for $H^{-1}(\cdot)$, $\psi_1(\cdot)$ and $\psi_2(\cdot)$, we adopt B-splines of order 3 on corresponding supports, with the number of interior knots being 0, 1 and 2 for $n = 250$, 500 and 1000 respectively. And the interior knots are picked to divide the corresponding support equally.

In Figure 1, we present the nonparametric estimation results. The first column of plots in Figure 1 depicts for different sample sizes the estimated curves of the cost function and the corresponding pointwise 95% confidence band, with lower and upper bounds being the 2.5% and 97.5% percentiles, respectively, based on bootstraps. And the second and third columns depict the estimated curve of disutility function under $W = \varpi_1$ and $W = \varpi_2$, respectively, and the corresponding pointwise 95% confidence bands. According to Figure 1, the estimated curves become closer to the true curves as the samples size increases. we also observe a clear pattern that all confidence bands shrink as the sample size increase. These indicate that our estimation procedure performs well, taking into consideration the nonparametric nature of our estimation procedures and that nonparametric procedures are generally known to be data-demanding.

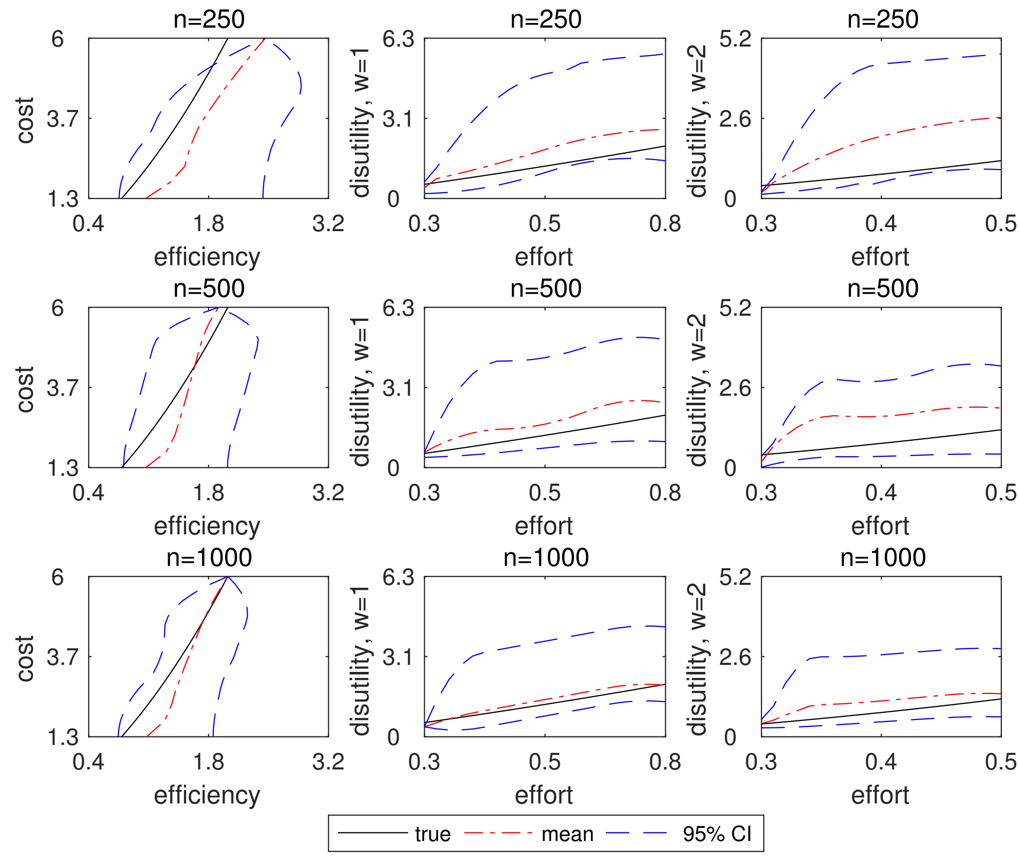
In Table 2 we report estimation results for γ (intertemporal weight) and α (bargaining power),[13] in terms of the mean, standard deviation (Std.), and root of mean squared error (RMSE). According to Table 2, as the sample size increases, the mean becomes closer to the true value, and both the Std. and RMSE decrease, suggesting good estimation performance.

7 Empirical illustrations

In this section, we apply our methods to analyze contracts for urban transportation services in France. The main objectives are to test the widely assumed LCC, and to evaluate how this specification assumption affects welfare assessment.

[13] As discussed, α and λ are not separately identifiable. Here we normalize $\lambda = 0.3$ and focus to estimate α.

Figure 1: Simulation results for the cost and disutility functions



Notes: The first column of plots report the simulation results of cost function. The second and third columns report that of disutility function under $W = \varpi_1$ and $W = \varpi_2$, respectively. In each plot, we depict in various dashed lines the estimated mean curve, and the corresponding 2.5%, 97.5% percentiles based on bootstraps. We also depict in solid lines the true curve.

7.1 Background and the data

The urban transportation industry in France is regulated. Since 1982, for each urban area with significant size and a public transportation network, a local authority (a city, a group of cities, or a district) must operate the network directly or contract with a single operator to provide passenger transportation service. In 2010, 91% of the urban public transportation contracts are implemented by operators, and only in few large cities such as Paris and Marseille, the transportation service is fully integrated within the city administration (Heddebaut, 2017).

In a French transportation contract, the local authority sets the bus route, fare structure, capacity, quality of service, etc., and the operator delivers the passenger transportation services specified in the contract. Motivated by private information of operators and limited auditing capacity of local authorities, we model the bilateral interaction between a local authority and an operator as a principal-agent problem, and assume away competition between agents.[14] In this industry, operators often have better information than local authorities about drivers' skills and behavior, fuel consumption, number of buses required for a certain route, and other cost-related factors. It also has been well documented that local authorities have difficulties to verify effort that operators put into cost-reducing and efficient management (e.g., see Domenach (1987)). Thus, the French urban transportation contracts fit our framework.

The dataset includes 543 two-period contracts implemented in France from 1987 to 2001, collected from the Centre d'Etude et de Recherche du Transport Urbain (CERTU, Lyon) surveys. For each contract in the dataset, the contract type (FF,CF, or CC), realized cost, subsidy in any FP period are recorded. Besides, additional characteristics of operators/contracts, such as local authorities' political ideology (whether the policymaker is left- or right-wing), the labor fee, number of employees, number of drivers, and size of rolling stock (measured by the number of vehicles), are recorded. Among these contracts, 281 are FF, 88 are CF, and the remaining 174 are CC. The same dataset is analyzed in Gagnepain et al. (2013) with a focus on the comparison between contracts with renegotiation and full commitment.

Table 3 provides summary statistics of the dataset. The average cost and subsidy are around

[14]Before 1993, the usual practice for local authorities was to simply award contracts with one operator via negotiation, and renewed the contracts after five years. In 1993, the 'Sapin' Act was promulgated, in hope of preventing corruption affairs and enhancing competition among operators. However, local authorities are neither obliged to choose the best (lowest) bid as in a standard procurement auction, nor required to state selection criteria in their documents. As a consequence, competition among operators plays minimal role in the French urban transportation industry, which is evidenced by the fact that, in more than 60% of auctions, only one operator submitted its bid, and for 90% of contracts the local authority renewed the contract with the incumbent instead of switching to another operator (Gautier and Yvrande-Billon, 2013). Therefore, following the existing literature (e.g., Gagnepain and Ivaldi (2002)), we assume away such competition, and apply the current principal-agent framework to the empirical study. Nevertheless, in our current framework, the parameter α, which measures the bargaining power of the operator, potentially captures competition, if there is any, among potential operators (Gagnepain et al., 2013). We believe that the simplified framework, together with the parameter α, provides a reasonable first-order approximation to the industry.

Table 3: Summary Statistics

Variables	Mean	Std. Dev.	Median	Min	Max
NO. of Contracts	543				
NO. of FF	281				
NO. of CF	88				
NO. of CC	174				
Cost	16860	15954	10347	2397	93993
Subsidy	18794	18236	12039	2265	114483
NO. of employees	413	364	267	68	1772
NO. of drivers	278	216	144	47	1182
Labor fee	10740	10241	6650	716	53178
Rolling stock	165	121	84	33	724
Right-wing	0.52	0.50	1.00	0.00	1.00

Notes: The units of cost, subsidy and labor fee are in 1000 euros.

17 million and 19 million euros, respectively, which suggests that operating the transportation network is on average profitable to the operators. The average labor fee is 10.7 million, accounting for 64 percent of the total cost. This suggests that reducing the labor fee is critical to increase the operator's profit. On average, 278 out of 413 (more than a half) employees are drivers, suggesting the transportation industry to be labor-intensive. As to the political ideology, over half of the contracts are signed with right-wing local authorities.

7.2 Testing for the LCC in the French transportation industry

First, we examine whether the LCC holds (i.e., whether operators' cost function is linear) for the French transportation industry. Specifically, we test the testable implication (3.9) of the LCC, by applying the inference procedure proposed in Section 3.2. Recall that the procedure requires a binary variable W, which is independent of firms' innate cost θ but potentially affects the disutility of effort.

Our choice of W is the political ideology dummy, which equals one if a local authority is right-wing, and zero otherwise. On the one hand, note that the innate cost θ captures factors that affect a firm's (in)efficiency but are inherently not under its control. In the transportation industry, these factors are mainly technologies related to urban-transportation that are accessible to an operator. Thus, it is reasonable to assume that the operator's accessibility to this kind of technologies has little to do with the local authority's political ideology. On the other hand, a firm can counterbalance a high innate cost by exerting managerial efforts, such as monitoring drivers, providing training programs (to promote driving habits that increase fuel efficiency), and solving potential conflicts, etc. The empirical evidence from Gagnepain and Ivaldi (2017) and Gagnepain and Ivaldi (2012) shows that right-wing local authorities care more about firms' efficiency and cost. Thus, right-wing authorities are more likely to accommodate operators' managerial efforts by providing some support, e.g., imposing less restrictive regulations, such that it could be less costly to exert efforts for operators who sign contracts with right-wing local authorities. Therefore, the political ideology dummy is a reasonable choice of the exclusion variable W for studying the French transportation industry.

Based on the FF and CF contracts in the data, the LCC test rejects the null hypothesis of a constant optimal effort at a significance level of 1%. As we discussed before, the constant optimal effort is a direct implication of the LCC, this testing result suggests that the LCC fails to adequately describe the cost structure in the French transportation industry. Instead, one should adopt a nonlinear specification for the cost function in analyzing the French transportation contracts, as what we do in the next section.

7.3 The welfare analysis

The second main objective of our empirical application is to conduct welfare analysis for the French transportation contracts. The assessment of welfare plays a central role in empirical studies of contracts. An example is the evaluation of welfare loss associated with renegotiation relative to full-commitment in Gagnepain et al. (2013). In this application, we evaluate how econometric specifications of cost function, i.e., LCC and non-LCC, affect welfare analysis of the contracts. For this purpose, we estimate welfare with and without imposing the LCC, then conduct a comparison. We first estimate the cost function, distribution of innate cost and other model primitives with (possibly) nonlinear cost, and estimate the same set of primitives under the LCC. Next, based on the estimation results, we conduct welfare analysis.

7.3.1 Estimation strategies

In the estimation, we use the share of drivers among all employees and the labor fee as X and Y, i.e., the two proxies of effort, respectively. Specifically, since employees mainly consist of drivers and engineers, X is constructed as

$$X = \frac{\#\text{ of drivers}}{\#\text{ of drivers} + \#\text{ of engineers}} = 1 - \text{ share of engineers.}$$

The choices of X and Y are plausible because: (i) The share of engineers provides a measure of endowed skills of an operator, thus is expected to be negatively related to the innate cost,[15] which in turn is positively related to the optimal effort. Therefore, we expect X to be negatively related to the optimal effort; (ii) Our choice for Y is in accordance with Cicala (2015)'s suggestion that one can employ cost-related variables to infer agent's effort. Specifically, the labor fee, which represents 64 percent of the total cost, can be reasonably interpreted as a function of the optimal

[15] As Gagnepain et al. (2013) point out, engineers are generally responsible for research and development, quality control, maintenance, and efficiency.

effort due to the one-to-one mapping between the optimal effort and the innate cost (under FP). We take into account the operator's rolling stock (i.e., number of transit vehicles) as a covariate Z that may affect firms' innate cost and effort directly, and affect cost and disutility functions indirectly. Recall that our identification strategies only identify the operator's bargaining power α and the cost of public funds λ up to the ratio $\alpha/(1+\lambda)$ (as in (2.16)). Nevertheless, for ease of discussion, we set $\lambda = 0.3$ as in Gagnepain et al. (2013), following the empirical suggestion by Ballard et al. (1985) that $\lambda \in [0.15, 0.40]$ in an efficient tax systems. This arrangement enables further identification and estimation of α.

Given that the LCC is rejected for the French transportation industry, the specification of firms' cost function is crucial for welfare analysis. In this application, we consider two alternative specifications of cost function: nonparametric and quadratic. A nonparametric specification imposes no functional form assumption on the cost function, and thus has the advantage of being more flexible. However, as for the disadvantages, the nonparametric specification is data-demanding, and causes difficulty in incorporating the covariate Z due to the issue of curse of dimensionality. A quadratic cost function simplifies estimation and is widely used in the literature on contracts, e.g., Luo et al. (2018b) discuss quadratic cost function in nonlinear pricing, Zhang (2021) employs a quadratic cost function in contracting with externalities. We estimate model primitives and conduct welfare assessments for both specifications of cost function. The disutility functions are nonparametric in both cases. For all involved nonparametric estimations, we adopt B-splines of order 3 on corresponding supports, with one interior knot.

It is challenging to incorporate the continous rolling stock Z into estimation, especially when the cost function is nonparametric. Because cost and disutility functions *implicitly* depend on Z through innate cost and optimal effort, it is difficult to model the dependence of cost on Z by using the existing nonparametric methods such as kernel estimation. Moreover, the sample size in this application is mild and the sample needs to be further categorized into two subsamples based on political ideology (W). Therefore, it is methodologically and practically infeasible to incorporate Z without imposing restrictive *ad hoc* assumptions. To address these issues, we take a simple and popular approach of discretizing Z, by assuming that Z plays a role in the model as two segments and estimating model primitives conditional on the two segments. Specifically, we discretize the rolling stock such that $Z = 1$ if the rolling stock is smaller than its sample mean, and $Z = 2$ otherwise.[16][17]

In the estimation, we obtain the two fixed prices for each of the four pairs of (W, Z) by averaging the fixed prices within the corresponding subsample.[18] Note that even though cost

[16] There are 343 and 200 contracts for $Z = 1$ and $Z = 2$, respectively. Among the 343 contracts with $Z = 1$, there are 121 and 83 FP ones with right-wing and left-wing, respectively. Among the 200 contracts with $Z = 2$, there are 89 and 76 FP ones with right-wing and left-wing, respectively.

[17] We test the LCC conditional on $Z = 1$ and $Z = 2$ and the results reject the LCC hypothesis at the 1% significance level.

[18] For contracts signed with right-wing local authorities, the normalized supports of effort are $[1, 15]$ and $[2, 30]$

and disutility functions depend on Z through innate cost and effort, their functional forms are specified as being the same for $Z = 1$ and $Z = 2$, as assumed in the exiting literature, e.g, Gagnepain et al. (2013). For all those parameters that explicitly depend on Z, we present the estimates for each Z, as well as their weighted average over Z.

In the case of LCC, we parameterize the cost function as $H(\theta - e) = \beta\,(\theta - e)$. As discussed in Section 3.2, a linear cost function induces a constant optimal effort under FP, regardless of θ. Thus, the two effort proxies X and Y are no longer necessary to identify and estimate the distribution of effort, because the distribution now degenerates to a single point (with probability mass one) for a give W. However, the lack of variation in optimal effort makes it impossible to identify the disutility functions $\{\psi_1(\cdot), \psi_2(\cdot)\}$ without further assumptions, as we show in the Appendix B.1. So we normalize $\psi_1(e) = e^2$, and adopt a simple parameterization $\psi_2(e) = \kappa e^2$ for ψ_2, which satisfies Assumption 3.2, and is widely used in the related literature, such as Laffont and Tirole (1988), Rogerson (2003), Chu and Sappington (2007), and Battaglini (2007).

7.3.2 Estimation results

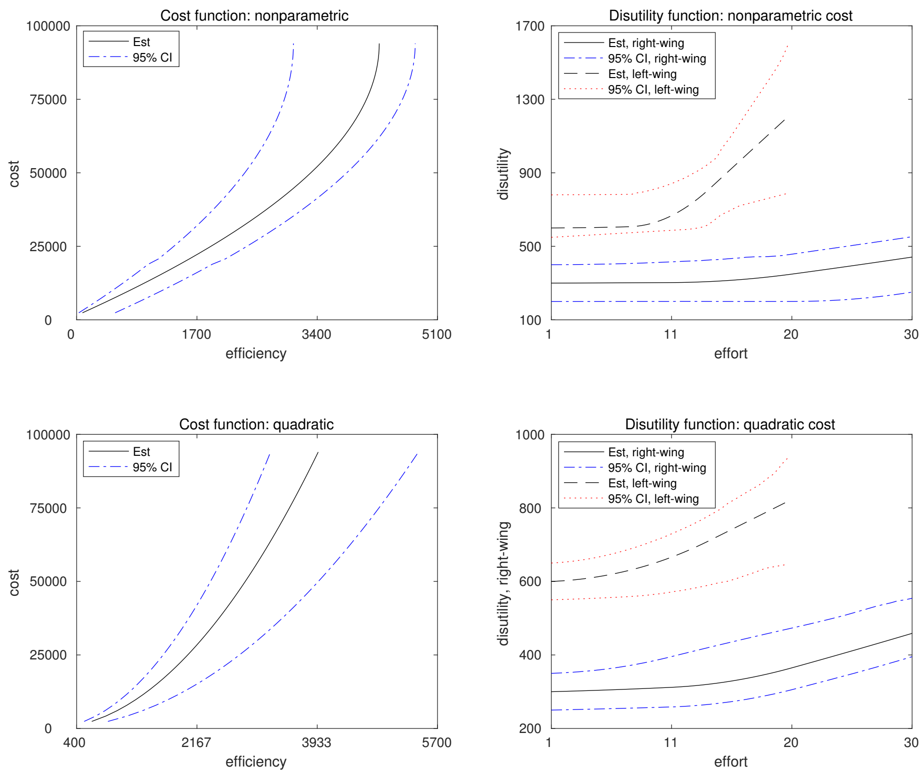
Figure 2 presents the estimated cost and disutility functions under the nonparametric and quadratic specifications of the cost function. According to Figure 2, for both the nonparametric and quadratic specifications, the estimated cost and marginal cost are both increasing in the efficiency. The estimated disutility function for left-wing authorities has a higher marginal disutility than that for right-wing authorities except at the lower bound of effort. This is consistent with our discussion on the choice of the variable W: It is less costly to exert managerial effort for operators who sign contracts with right-wing local authorities. Moreover, Figure 2 illustrates that the estimated cost and disutility functions under the quadratic specification of cost are similar to those under the nonparametric specification.

The top panel of Table 4 reports estimates of other model parameters under the nonparametric cost function. The estimated intertemporal weight γ is 0.389, which suggests that operators pay relatively more attention to the second-period profit than the first-period profit. The estimated operator's bargaining power α is 0.830, which is consistent with the theoretical restriction that $\alpha < 1 + \lambda$. The middle panel of Table 4 presents all the estimated parameters when cost is specified as a quadratic function. The estimated coefficients β_1 and β_2 are used to plot the bottom-left plot in Figure 2. The parameters γ and α are estimated to be 0.340 and 0.936, respectively.

The estimation results under the LCC are reported in the bottom panel of Table 4. A comparison of estimates of γ and α across three panels in the table shows that the estimates under the quadratic and nonparametric specifications of cost are very close to each other, and

for $Z = 1$ and $Z = 2$, respectively. Similarly, for left-wing, the supports are $[1, 10]$ and $[2, 20]$ for $Z = 1$ and $Z = 2$, respectively. Note that for a given Z, the lower bounds of effort for right-wing and left-wing are the same, which is consistent with our identification argument.

Figure 2: Estimation results of cost and disutility functions



Notes: The left two plots report the estimation results of cost function, and the right two report the estimation results of disutility function for the right-wing and left-wing. We use $[\min\{c_i\}_{i=1}^n, \max\{c_i\}_{i=1}^n]$ as the cost support. The normalized supports of the effort are $[1, 20]$ for the left wing and $[1, 30]$ for the right wing. In each of the left two panels, the solid line depicts the estimated curve, and the two dashed lines indicate the 2.5% and 97.5% percentiles (corresponding to the 95% confidence interval (CI)). In each of the right two panels, the solid and dashed lines depict the estimated curves for the right-wing and the left-wing, respectively. The two dash-dotted lines and the two dotted lines indicate the 2.5% and 97.5% percentiles (corresponding to the 95% CI) for the right-wing and the left-wing, respectively. All percentiles are based on 500 bootstrap replications.

Table 4: Estimates of parameters with quadratic and nonparametric cost functions

nonparametric cost function								
Parameter			Intertemporal weight γ			Bargaining power α		
			$z=1$	$z=2$	Average	$z=1$	$z=2$	Average
Estimate			0.450***	0.283***	0.389***	0.899**	0.711***	0.830***
			(0.083)	(0.071)	(0.062)	(0.370)	(0.182)	(0.242)
quadratic cost function								
Parameter	Cost		Intertemporal weight γ			Bargaining power α		
	β_1	β_2	$z=1$	$z=2$	Average	$z=1$	$z=2$	Average
Estimate	0.125***	0.006***	0.387***	0.260***	0.340***	0.970***	0.877***	0.936***
	(0.043)	(0.002)	(0.089)	(0.075)	(0.062)	(0.106)	(0.299)	(0.125)
linear cost function								
Parameter	Cost	Disutility	Intertemporal weight γ			Bargaining power α		
	β	κ	$z=1$	$z=2$	Average	$z=1$	$z=2$	Average
Estimate	56.945**	0.232**	0.960***	0.416	0.760***	1.209***	1.201***	1.206***
	(16.783)	(0.105)	(0.340)	(0.369)	(0.254)	(0.221)	(0.393)	(0.198)

Notes: The top, middle, and bottom panels present parameter estimates under nonparametric, quadratic and linear cost functions, respectively. The quadratic cost function is specified as $H(\theta-e)=\beta_1(\theta-e)+\beta_2(\theta-e)^2$ and the linear cost is $H(\theta-e)=\beta(\theta-e)$. For each value of Z, parameters γ and α are first estimated conditional on W, the estimates reported in the table are averages over W. Th column "Average" is the weighted average for $Z=1$ and $Z=2$ with the weights being their respective frequencies. Standard errors in parentheses are obtained by bootstrapping 500 times. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

they are both drastically different from the estimates under the LCC. An implication is that a quadratic function could be a reasonable approximation of the operators' costs for the French transportation industry.

7.3.3 Welfare assessments and comparison

Now we evaluate how the specification of cost function affects the welfare assessment of the contracts based on the estimates of parameters. For the evaluation, we focus on the average social welfare, $SW \equiv S-(1+\lambda)\,Q+\alpha U$. The assessment of SW is delicate in the sense that SW consists of several components, each of which needs to be estimated separately and then put together. The details on the welfare assessment are provided in Appendix B.2.

We report the welfare assessments and comparison results in Table 5. As indicated, whether or not the LCC is imposed makes a substantial difference in assessing the average social welfare. For the French transportation industry, LCC would overestimate the welfare about 43 million euros relative to the case with a nonparametric cost function, which is about 2.5 times of the average contract cost (16.86 million euros by Table 1). It also worth noting that the fully nonparametric specification and the quadratic specification lead to very similar assessments of welfare. These findings suggest that the welfare assessments are very sensitive to the cost function being specified as LCC or not. Thus, one needs to be cautious on deciding whether or not to impose linearity, since mis-specification could lead to substantial bias. In addition, we note that these findings are reminiscent of An and Zhang (2018), who show in theory that, for single-period contracts, the social welfare can differ a lot depending on whether the cost function is linear or not.

Table 5: Assessments of welfare under different cost specifications

Welfare items	Model specification	Estimate		
		$Z=1$	$Z=2$	Average
Social cost $(1+\lambda)\,Q$	Nonparametric cost	12.720***	52.683***	27.439***
		(0.844)	(2.641)	(1.332)
	Quadratic cost	11.080***	43.466***	23.008***
		(0.716)	(0.644)	(3.530)
	Linear cost	9.385***	27.810***	16.171***
		(1.056)	(4.330)	(1.843)
Informational rent αU	Nonparametric cost	5.455*	19.587**	10.660***
		(3.193)	(7.814)	(3.262)
	Quadratic cost	1.644	4.689***	2.765***
		(2.269)	(1.237)	(0.931)
	Linear cost	26.163***	71.608***	42.901***
		(0.832)	(5.430)	(2.027)
Welfare difference $\alpha\Delta U-(1+\lambda)\Delta Q$	Nonpara. vs. Linear	-24.043	-76.893	-43.509
	Quad. vs. Linear	-26.214	-82.575	-46.973
	Nonpara. vs. Quad.	2.171	5.682	3.464

Notes: All estimates are in million euros. The welfare terms are first estimated conditional on $\{W, Z\}$. Columns "$Z = 1$" and "$Z = 2$" report their weighted averages over W (but still conditional on Z), and the column "Average" reports their weighted averages over both W and Z, with the weights being W's and Z's frequencies of taking different values. Standard errors in parentheses are bootstrapped 500 times. *, ** and *** indicate the corresponding p-value $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

8 Conclusion

We provide a rigorous econometric framework for analyzing FP-CR contracts in a two-period settings. We first present a rigorous testing procedure to test the widely assumed linear cost assumption in the literature of contracts. Next, we prove that the simple contract model without parametric assumptions on the cost and disutility functions are nonparametrically identified on intervals corresponding to FF and CF contracts. The method of identification is applicable to a large class of simple contracts. Following closely the identification strategy, we propose a nonparametric procedure to estimate the model primitives. The proposed testing and estimation procedures are shown to have good finite-sample performance in Monte Carlo simulations. In the empirical application, we apply our methodology to the public transportation procurement contracts in France. We find strong evidence against linearity of the cost function. By comparing welfare across different model specifications, we illustrated that imposing the LCC incorrectly would seriously bias the assessments of contracts, while a fully nonparametric specification and a parametric nonlinear specification lead to very similar assessments of welfare.

Our identification strategy can be potentially extended to handle linear cost sharing (LCS) contracts, which encompass FP and CR as special cases. Intuitively, observing the more flexible payment from LCS contracts (compared with that of FP or CR contracts) would provide additional information that is rich enough to identify the additional finite dimensional parameter of LCS contracts.

References

Aaker, D. and J. Myers, *Advertising management* (Prentice-Hall, 1987). 1

Abito, J. M., "Agency Costs in Environmental Regulation: Evidence from Regulated Electric Utilities," Working Paper, University of Pennsylvania, 2014. 1

Ai, C. and X. Chen, "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica* 71 (2003), 1795–1843. 5

An, Y. and D. Zhang, "Simple Menus of Cost-based Contracts with Monotone Optimal Efforts," *Journal of Public Economic Theory* 20 (2018), 742–756. 7.3.3

Angrist, J., V. Chernozhukov and I. Fernández-Val, "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica* 74 (2006), 539–563. 3.2, C

Aryal, G., I. Perrigne and Q. Vuong, "Identification of Insurance Models with Multidimensional Screening," Working paper, Australian National University, 2010. 1

Bajari, P. and S. Tadelis, "Incentives versus transaction costs: A theory of procurement contracts," *RAND Journal of Economics* 32 (2001), 387–407. 1

Ballard, C. L., J. B. Shoven and J. Whalley, "General equilibrium computations of the marginal welfare costs of taxes in the United States," *American Economic Review* 75 (1985), 128–138. 7.3.1

Banerjee, A. V. and E. Duflo, "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry," *Quarterly Journal of Economics* 115 (2000), 989–1017. 1

Baron and Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica* 50 (1982), 911–930. 2.1, 2

Baron, D. P., "Regulation and legislative choice," *RAND Journal of Economics* 19 (1988), 467–477. 2

Battaglini, M., "Optimality and renegotiation in dynamic contracting," *Games and economic behavior* 60 (2007), 213–246. 7.3.1, B.1

Carroll, R. J., X. Chen and Y. Hu, "Identification and estimation of nonlinear models using two samples with nonclassical measurement errors," *Journal of nonparametric statistics* 22 (2010), 379–399. 5

Chen, X., "Large sample sieve estimation of semi-nonparametric models," *Handbook of econometrics* 6 (2007), 5549–5632. 5

Chen, X., Z. Liao and Y. Sun, "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics* 178 (2014), 639–658. 5

Chernozhukov, V. and C. Hansen, "An IV model of quantile treatment effects," *Econometrica* 73 (2005), 245–261. 4.2

Chesher, A., "Identification in nonseparable models," *Econometrica* 71 (2003), 1405–1441. 4.2

Chu, L. Y. and D. E. Sappington, "Simple cost-sharing contracts," *American Economic Review* 97 (2007), 419–428. 7.3.1, B.1

Cicala, S., "When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation," *American Economic Review* 105 (2015), 411–444. 4.1, 7.3.1

D'Haultfoeuille, X. and P. Février, "The provision of wage incentives: A structural estimation using contracts variation," Working paper, CREST-ENSAE, 2019. 1, 4.2

Domenach, O., "Métamorphose des rapports contractuels dans les transports publics urbains: de l'autorité concédante à l'autorité organisatrice et du concessionnaire au prestataire de service," *Transports* 32 (1987), 112–120. 7.1

Freyberger, J. and M. A. Masten, "A practical guide to compact infinite dimensional parameter spaces," *Econometric Reviews* 38 (2019), 979–1006. A

Gagnepain, P. and M. Ivaldi, "Incentive regulatory policies: the case of public transit systems in France," *RAND Journal of Economics* 33 (2002), 605–629. 1, 14

———, "Choice of regulation, incentives, and political capture: The case of local transport services in France," Technical Report, CEPR working Paper, 2012. 7.2

———, "Economic efficiency and political capture in public service contracts," *The Journal of Industrial Economics* 65 (2017), 1–38. 7.2

Gagnepain, P., M. Ivaldi and D. Martimort, "The Cost of Contract Renegotiation: Evidence from the Local Public Sector," *American Economic Review* 103 (2013), 2352–83. 1, 2, 2.3, 2.3, 7.1, 14, 7.3, 15, 7.3.1, A, C

Gautier, A. and A. Yvrande-Billon, "Contract renewal as an incentive device. An application to the French urban public transport sector," *Review of Economics and Institutions* 4 (2013), 29. 14

Gopal, A., K. Sivaramakrishnan, M. S. Krishnan and T. Mukhopadhyay, "Contracts in offshore software development: An empirical analysis," *Management Science* 49 (2003), 1671–1683. 1

Guerre, E., I. Perrigne and Q. Vuong, "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica* 77 (2009), 1193–1227. 4.2

Heckman, J. J., R. L. Matzkin and L. Nesheim, "Nonparametric identification and estimation of nonadditive hedonic models," *Econometrica* 78 (2010), 1569–1591. 4.2

Heddebaut, O., *The evolution of public transport contracts in France*, Ph.D. thesis, IFSTTAR-Institut Français des Sciences et Technologies des Transports (2017). 7.1

Hu, Y., "The Econometrics of Unobservables: Applications of Measurement Error Models in Empirical Industrial Rrganization and Labor Economics," *Journal of Econometrics* 200 (2017), 154–168. 1

Laffont, J.-J. and J. Tirole, "Using cost observation to regulate firms," *Journal of Political Economy* 94 (1986), 614–641. 1

———, "The dynamics of incentive contracts," *Econometrica* (1988), 1153–1175. 7.3.1, B.1

———, "Adverse selection and renegotiation in procurement," *Review of Economic Studies* 57 (1990), 597–625. 2.3, 2.3

———, *A Theory of Incentives in Procurement and Regulation* (MIT Press, 1993). 2.1

Li, T., "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics* 110 (2002), 1–26. 4.1

Luo, Y., I. Perrigne and Q. Vuong, "Structural Analysis of Nonlinear Pricing," *forthcoming, Journal of Political Economy* (2018a). 1, 4.4

———, “Structural analysis of nonlinear pricing,” *Journal of Political Economy* 126 (2018b), 2523–2568. 7.3.1

Perrigne, I. and Q. Vuong, “Nonparametric identification of a contract model with adverse selection and moral hazard,” *Econometrica* 79 (2011), 1499–1539. 1, 8

———, “On the identification of the procurement model,” *Economics Letters* 114 (2012), 9–11. 8

Rogerson, W. P., “An economic framework for analyzing DoD profit policy,” Technical Report, RAND Corp Santa Monica, CA, 1992. 1

———, “Simple menus of contracts in cost-based procurement and regulation,” *American Economic Review* 93 (2003), 919–926. 1, 7.3.1, B.1

Schennach, S. M. and Y. Hu, “Nonparametric identification and semiparametric estimation of classical measurement error models without side information,” *Journal of the American Statistical Association* 108 (2013), 177–186. 1, 4.1, 4.1, D.2.2

Shen, X., “On methods of sieves and penalization,” *The Annals of Statistics* (1997), 2555–2591. 5, 5

Torgovitsky, A., “Identification of nonseparable models using instruments with small support,” *Econometrica* 83 (2015), 1185–1197. 4.2

van der Vaart, A., *Asymptotic Statistics* (Cambridge University Press, 1999). A

van der Vaart, A. W. and J. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics* (Springer, New York, 1996). A.3, A, A, C, C

Zhang, D., “Testing Passive versus Symmetric Beliefs in Contracting with Externalities,” *International Economic Review* 62 (2021), 723–767. 7.3.1

APPENDIX

The appendix is organized as follows: In Appendix A, we provide proofs of the main results. In Appendix B, we provide additional details on the empirical study. In Appendix C, we provide proofs of Lemmas A.1 - A.3, which are employed (in Appendix A) for proving the main results. Appendix D collects additional supplementary contents.

A Proofs of the main results

We present proofs of main results in the order of their occurrences in the main text: Lemma 2.1, Propositions 2.1, 2.2, Corollary 3.1, Theorem 1, Lemma 4.1, Propositions 4.1, 4.2, Theorem 2 and Proposition 5.1. Note that Corollary 3.1 simply restates the results in Lemma 2.1 and Proposition 2.2, but conditioned on W. Also note that Theorem 2 simply summarizes identification results stated in Proposition 4.2 and Section 4.3-4.4, all of which are proven/explained elsewhere. So dedicated proofs for Corollary 3.1 and Theorem 2 are unnecessary.

In proving the main results, we utilize the following technical lemmas, labeled as Lemmas A.1 – A.3, whose proofs are provided in Appendix C.

Lemma A.1 (Renegotiation-proof) *There is no loss of generality in restricting the analysis to contracts of the form* $\mathbf{b} = (b_1^{FF}, b_2^{FF}, b_2^{CF})$ *that come unchanged through the renegotiation process, that is,* $R \equiv (b_2^{FF}, b_2^{CF})$ *maximizes the principal's second-period welfare subject to the following acceptance conditions:*

$$\tilde{b}_2^{FF} \geq b_2^{FF} \quad \text{and} \quad \tilde{b}_2^{CF} \geq b_2^{CF}. \tag{A.1}$$

Lemma A.2 *An initial offer* $\mathbf{b} = \left(b_1^{FF}, b_2^{FF}, b_2^{CF}\right)$ *is renegotiation-proof if and only if the following two conditions hold:*

$$\theta_2\left(b_2^{CF}\right) \geq \theta_1\left(\mathbf{b}\right); \tag{A.2}$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right) \frac{F\left(\theta_2\left(b_2^{CF}\right)\right) - F\left(\theta_1\left(\mathbf{b}\right)\right)}{f\left(\theta_2\left(b_2^{CF}\right)\right)} \quad \geq \quad \frac{\left[H\left(\theta_2\left(b_2^{CF}\right)\right) - b_2^{CF}\right]}{H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right)}, \tag{A.3}$$

where $\theta_1\left(\mathbf{b}\right)$ *is the cutoff type between FF and CF induced by* C*, and* $\theta_2\left(b_2^{CF}\right)$ *is the cutoff one between CF and CC induced by* C*. Note that, as its notation suggests,* $\theta_2\left(b_2^{CF}\right)$ *depends on* b_2^{CF}*, but not* b_1^{FF} *or* b_2^{FF}*, under renegotiation-proof.*

Lemma A.3 (Uniform weak convergence) *Let Assumption 3.1, 3.2, and Condition T.1 hold. It holds that*

$$\begin{aligned}
&\sqrt{n_f}\left\{\left[\widehat{C}_t\left(\tau;\varpi_1\right) - \widehat{C}_t\left(\tau p;\varpi_2\right)\right] - \left[\widehat{C}_t\left(\nu;\varpi_1\right) - \widehat{C}_t\left(\nu p;\varpi_2\right)\right]\right\} \\
&- \sqrt{n_f}\left\{\left[C_t\left(\tau;\varpi_1\right) - C_t\left(\tau p;\varpi_2\right)\right] - \left[C_t\left(\nu;\varpi_1\right) - C_t\left(\nu p;\varpi_2\right)\right]\right\} \\
\xrightarrow{\mathcal{L}} \;& (1,0)\, z\left(\tau\right)/J\left(\tau\right) - (1,1)\, z\left(\tau p\right)/J\left(\tau p\right) - (1,0)\, z\left(\nu\right)/J\left(\nu\right) + (1,1)\, z\left(\nu p\right)/J\left(\nu p\right)
\end{aligned}$$

where: (i) $J(\tau) \equiv \mathbb{E}\left[f_{C_t|W}\left(q_1(\tau) + q_2(\tau)\mathbf{1}(W = \varpi_2)\,|\varpi_2\right)\mathbf{1}(W = \varpi_2)\right]$ *for* $\tau \in [0,1]$*, with*

$$(q_1(\tau), q_2(\tau)) \equiv \underset{q_1, q_2}{\operatorname{argmin}}\, \mathbb{E}\left[\rho_\tau\left(C - q_1 - q_2 \cdot \mathbf{1}(W = \varpi_2)\right)|\textit{FP for period-}t\right];$$

(ii) $z(\cdot)$ *is a zero mean tight Gaussian process on* $L^\infty[0,1]$*; (iii) The weak convergence "*$\xrightarrow{\mathcal{L}}$*" is in the sense of Chapter 1.3 in van der Vaart and Wellner (1996).*

Proof of Lemma 2.1. As discussed in the main context, since the CR option provides no incentive to exert effort, the optimal effort is always zero under CR.

Taking first-order derivative w.r.t. θ on both side of the f.o.c. (2.2) (i.e., $H'(\theta - e(\theta)) = \psi'(e(\theta))$) yields

$$e'(\theta) = \frac{H''(\theta - e(\theta))}{H''(\theta - e(\theta)) + \psi''(e(\theta))} \in (0,1), \tag{A.4}$$

where the strict inequalities follow from the conditions $H''(\cdot) > 0$ and $\psi''(\cdot) > 0$ by Assumption 2.1. (A.4) verifies the second claim.

It follows from (A.4) and $H''(\cdot) > 0$ that

$$\begin{aligned}\frac{d[\theta - e(\theta)]}{d\theta} &= 1 - e'(\theta) > 0,\\ \frac{dH(\theta - e(\theta))}{d\theta} &= H''(\theta - e(\theta))\left[1 - e'(\theta)\right] > 0,\end{aligned}$$

which verifies the third claim. ■

Proof of Proposition 2.1. For any given fixed price b within suitable range,[19] there exists a unique cutoff value $\theta(b)$ s.t. an agent with $\theta = \theta(b)$ is indifferent between FP (with fixed price b) and CR, i.e., these two options lead to the same amount of payoff for him. Recall from (2.3) that the payoff under FP is

$$u_F(\theta, b) = b - H(\theta - e(\theta)) - \psi(e(\theta)).$$

Also recall that, under the normalizing condition $\psi(0) = 0$ by Assumption 2.1, the payoff under CR would always be zero. Therefore, we have

$$u_F(\theta(b), b) = b - H(\theta(b) - e(\theta(b))) - \psi(e(\theta(b))) = 0. \tag{A.5}$$

As a special case of (A.5) above, we have $u_F(\theta^*, b^*) = b^* - H(\theta^* - e(\theta^*)) - \psi(e(\theta^*)) = 0$, with $\theta^* \equiv \theta(b^*)$, which verifies (2.8) in Proposition 2.1 (ii).

[19] Specifically, $b \in \left[H(\underline{\theta} - e(\underline{\theta})) + \psi(e(\underline{\theta})), H(\bar{\theta} - e(\bar{\theta})) + \psi(e(\bar{\theta}))\right]$. Note that, if b is below this range, CR would be preferred by all θ. Similarly, if b is above this range, FP would be preferred by all θ.

To verify Proposition 2.1 (i), note that, for a given b, we have

$$\frac{\partial u_F(\theta,b)}{\partial\theta} = -\frac{dH(\theta-e(\theta))}{d\theta} - \psi^{'}(e(\theta))\cdot\frac{de(\theta)}{d\theta} < 0,$$

where the inequality follows from that $\frac{dH(\theta-e(\theta))}{d\theta} > 0$ and $\frac{de(\theta)}{d\theta} > 0$ (according to Lemma 2.1) and that $\psi^{'}(\cdot) > 0$ (by Assumption 2.1 (ii)). Consequently, $u_F(\theta,b) > u_F(\theta(b),b) = 0$ for all $\theta < \theta(b)$, and $u_F(\theta,b) < u_F(\theta(b),b) = 0$ for all $\theta > \theta(b)$. In particular, for $b = b^*$, we have $u_F(\theta,b^*) > u_F(\theta^*,b^*) = 0$ for all $\theta < \theta^*$, and $u_F(\theta,b^*) < u_F(\theta^*,b^*) = 0$ for all $\theta > \theta^*$, which verify Proposition 2.1 (i).

We verify (2.9) in Proposition 2.1 (ii) as follows: According to (A.5), we can rewrite $\tilde{\pi}(\theta,b)$ (specified by (2.6)) as

$$\tilde{\pi}(\theta,b) = \begin{cases} S-(1+\lambda)b+\alpha[b-H(\theta-e(\theta))-\psi(e(\theta))], & \text{for } \theta\le\theta(b); \\ S-(1+\lambda)H(\theta), & \text{for } \theta>\theta(b). \end{cases}$$

In turn, we can rewrite $\pi(b)$ (specified by (2.7)) as

$$\begin{aligned} \pi(b) &= \int_{\underline{\theta}}^{\bar{\theta}}\tilde{\pi}(\theta,b)\,dF(\theta) \\ &= S-(1+\lambda)F(\theta(b))b+\alpha F(\theta(b))b+\alpha\int_{\underline{\theta}}^{\theta(b)}[H(\theta-e(\theta))+\psi(e(\theta))]\,dF(\theta) \\ &\quad -(1+\lambda)\int_{\theta(b)}^{\bar{\theta}}H(\theta)\,dF(\theta). \end{aligned}$$

Since $b^* = \underset{b}{\operatorname{argmax}}\,\pi(b)$, the following f.o.c. holds:

$$\frac{d\pi(b^*)}{db} = [\alpha-(1+\lambda)]F(\theta^*)-(1+\lambda)f(\theta^*)\theta^{'}(b^*)b^*+(1+\lambda)H(\theta^*)f(\theta^*)\theta^{'}(b^*) = 0$$

for which rearranging terms yields

$$\left(1-\frac{\alpha}{1+\lambda}\right)\frac{F(\theta^*)}{f(\theta^*)} = [H(\theta^*)-b^*]\,\theta^{'}(b^*). \tag{A.6}$$

Note that taking derivative w.r.t. b on both side of (A.5) and rearranging terms yield

$$\begin{aligned} \theta^{'}(b) &= \left\{H^{'}(\theta(b)-e(\theta(b)))-\left[H^{'}(\theta(b)-e(\theta(b)))-\psi^{'}(e(\theta(b)))\right]e^{'}(\theta(b))\right\}^{-1} \\ &= \frac{1}{H^{'}(\theta(b)-e(\theta(b)))}, \end{aligned}$$

where the second equality is due to that $H^{'}(\theta(b)-e(\theta(b)))-\psi^{'}(e(\theta(b))) = 0$ according to (2.2). As a special case of the equation above, for $b = b^*$, we have

$$\theta^{'}(b^*) = H^{'}[(\theta^*-e(\theta^*))]^{-1}. \tag{A.7}$$

Substituting (A.7) into (A.6) yields

$$\left(1-\frac{\alpha}{1+\lambda}\right)\frac{F\left(\theta^*\right)}{f\left(\theta^*\right)} \quad = \quad \frac{H\left(\theta^*\right)-b^*}{H'\left(\theta^*-e\left(\theta^*\right)\right)},$$

which completes the proof. ■

Proof of Proposition 2.2.

According to (2.10) – (2.12), the expected intertemporal welfare when offering $\mathbf{b}=\left(b_1^{FF}, b_2^{FF}, b_2^{CF}\right)$ is

$$\begin{aligned}
\pi_{int}\left(\mathbf{b}\right) \quad = \quad & \int_{\underline{\theta}}^{\theta_1(\mathbf{b})}\left\{S-(1+\lambda)(\gamma b_1^{FF}+(1-\gamma)b_2^{FF})\right.\\
& \left.+\,\alpha\left[\gamma b_1^{FF}+(1-\gamma)b_2^{FF}-H(\theta-e(\theta))-\psi(e(\theta))\right]\right\}dF(\theta)\\
& +\int_{\theta_1(\mathbf{b})}^{\theta_2\left(b_2^{CF}\right)}\left\{S-(1+\lambda)(\gamma H(\theta)+(1-\gamma)b_2^{CF})\right.\\
& \left.+\,\alpha(1-\gamma)\left[b_2^{CF}-H(\theta-e(\theta))-\psi(e(\theta))\right]\right\}dF(\theta)\\
& +\int_{\theta_2\left(b_2^{CF}\right)}^{\bar{\theta}}\left[S-(1+\lambda)H(\theta)\right]dF(\theta).
\end{aligned} \tag{A.8}$$

The optimal renegotiation-proof menu solves the following optimization problem:[20]

$$\max_{\mathbf{b}}\ \pi_{int}(\mathbf{b}) \quad \text{subject to (A.3)}. \tag{A.9}$$

According to Lemma A.2, the optimal fixed prices $\mathbf{b}^\dagger=\left(b_1^{FF\dagger}, b_2^{FF\dagger}, b_2^{CF\dagger}\right)$ and the corresponding cutoff types $(\theta_1\left(\mathbf{b}^\dagger\right), \theta_2\left(b_2^{CF\dagger}\right))$ satisfy

$$b_2^{CF\dagger}=H\left(\theta_2\left(b_2^{CF\dagger}\right)-e\left(\theta_2\left(b_2^{CF\dagger}\right)\right)\right)+\psi\left(e\left(\theta_2\left(b_2^{CF\dagger}\right)\right)\right), \tag{A.10}$$

$$b_1^{FF\dagger}+\frac{(1-\gamma)}{\gamma}\left(b_2^{FF\dagger}-b_2^{CF\dagger}\right)=H\left(\theta_1\left(\mathbf{b}^\dagger\right)-e\left(\theta_1\left(\mathbf{b}^\dagger\right)\right)\right)+\psi\left(e\left(\theta_1\left(\mathbf{b}^\dagger\right)\right)\right). \tag{A.11}$$

Due to the fact that $\partial\theta_1\left(\mathbf{b}^\dagger\right)/\partial b_1^{FF}=\left[\gamma/(1-\gamma)\right]\partial\theta_1\left(\mathbf{b}^\dagger\right)\partial b_2^{FF}$ by (A.11), it can be shown that the f.o.c.'s for $b_1^{FF\dagger}$ and $b_2^{FF\dagger}$ are identical, and hence lead to the same optimal solution $b_1^{FF\dagger}=b_2^{FF\dagger}\equiv\underline{b}$.

Next, define $\bar{b}\equiv b_2^{CF\dagger}$. For expositional simplicity, we suppress the dependences of θ_1 and θ_2 on $(\underline{b},\bar{b})$ and $\bar{b}$ respectively. Following similar arguments as in the discussion of Proposition 2 in Gagnepain et al. (2013), the renegotiation-proof constraint (A.3) is binding in equilibrium, i.e., we have

$$\left(1-\frac{\alpha}{1+\lambda}\right)\frac{F\left(\theta_2\left(b_2^{CF}\right)\right)-F\left(\theta_1\left(\mathbf{b}\right)\right)}{f\left(\theta_2\left(b_2^{CF}\right)\right)} \quad = \quad \frac{\left[H\left(\theta_2\left(b_2^{CF}\right)\right)-b_2^{CF}\right]}{H'\left(\theta_2\left(b_2^{CF}\right)-e\left(\theta_2\left(b_2^{CF}\right)\right)\right)}. \tag{A.12}$$

[20] We assume (A.2) holds with strict inequality.

Note that, when plugging the optimal $\left(\underline{b}, \underline{b}, \bar{b}\right)'$ and $\bar{b}$ into $\theta_1(\cdot)$ and $\theta_2(\cdot)$, respectively, (A.10) - (A.12) immediately imply (2.14) - (2.16) of Proposition 2.2.

For simplicity of presentation, hereafter in the proof of Proposition 2.2, we suppress the dependence of $\theta_1(\cdot)$ and $\theta_2(\cdot)$ on $\mathbf{b}$ by using simplified notations $\theta_1 \equiv \theta_1(\mathbf{b})$ and $\theta_2 \equiv \theta_2\left(b_2^{CF}\right)$. Next, we show that the following two equations hold in equilibrium:

$$\left(1-\frac{\alpha}{1+\lambda}\right)\left[F(\theta_2)-F(\theta_1)\right]=\frac{f(\theta_2)\, s_2(\theta_2)}{H'(\theta_2-e(\theta_2))}-\frac{f(\theta_1)\, s_1(\theta_1)}{H'(\theta_1-e(\theta_1))}+\frac{\vartheta\cdot s(\theta_1,\theta_2)}{(1+\lambda)(1-\gamma)}, \quad \text{(A.13)}$$

$$\frac{\gamma(1+\lambda-\alpha)F(\theta_1)}{f(\theta_1)}=\frac{\gamma(1+\lambda)s_1(\theta_1)+\vartheta\left(1-\frac{\alpha}{1+\lambda}\right)}{H'(\theta_1-e(\theta_1))}, \quad \text{(A.14)}$$

where

$$\begin{aligned}
s_1(\theta_1) &\equiv H(\theta_1)-H(\theta_1-e(\theta_1))-\psi(e(\theta_1)),\\
s_2(\theta_2) &\equiv H(\theta_2)-H(\theta_2-e(\theta_2))-\psi(e(\theta_2)),\\
s(\theta_1,\theta_2) &\equiv \left(\frac{\alpha}{1+\lambda}-1\right)\left(\frac{f(\theta_2)}{H'(\theta_2-e(\theta_2))}-\frac{(\gamma-1)}{\gamma}\frac{f(\theta_1)}{H'(\theta_1-e(\theta_1))}\right)\\
&\quad+\frac{\left[H'(\theta_2)-H'(\theta_2-e(\theta_2))\right]f(\theta_2)+f'(\theta_2)\,s_2(\theta_2)}{\left[H'(\theta_2-e(\theta_2))\right]^2}\\
&\quad-\frac{H''\left[\theta_2-e(\theta_2)\right]\left[1-e'(\theta_2)\right]f(\theta_2)\,s_2(\theta_2)}{\left[H'(\theta_2-e(\theta_2))\right]^3},
\end{aligned}$$

and $\vartheta > 0$ is the Lagrange multiplier corresponding to the binding constraint (A.12).

(A.14) and (A.13) can be derived as follows: Using (2.2), we take the first derivatives of θ_1 in (A.11) with respect to $\underline{b}$ and $\bar{b}$ to obtain

$$\frac{d\theta_1}{d\underline{b}}=\frac{1}{\gamma}\frac{1}{H'(\theta_1-e(\theta_1))} \text{ and } \frac{d\theta_1}{d\bar{b}}=\frac{\gamma-1}{\gamma}\frac{1}{H'(\theta_1-e(\theta_1))}. \quad \text{(A.15)}$$

Similarly, taking the first derivatives of θ_2 in (A.10) with respect to $\underline{b}$ and $\bar{b}$ implies

$$\frac{d\theta_2}{d\underline{b}}=0 \text{ and } \frac{d\theta_2}{d\bar{b}}=\frac{1}{H'(\theta_2-e(\theta_2))}. \quad \text{(A.16)}$$

Let $\mathcal{L}\left(\underline{b},\bar{b}\right)=\pi_{int}\left(\underline{b},\bar{b}\right)+\vartheta\cdot s_0\left(\underline{b},\bar{b}\right)$, with $s_0\left(\underline{b},\bar{b}\right)$ defined as

$$s_0\left(\underline{b},\bar{b}\right)\equiv\frac{\left[H(\theta_2)-b_3\right]f(\theta_2)}{H'\left(\underline{\theta}-e(\theta_2)\right)}-\left(1-\frac{\alpha}{1+\lambda}\right)\left[F(\theta_2)-F(\theta_1)\right]\leq 0, \quad \text{(A.17)}$$

corresponding to constraint (A.12). Using (A.11), (A.10), (A.15), and (A.16), we can show that

$$\frac{\partial \pi_{int}\left(\underline{b},\bar{b}\right)}{\partial \underline{b}} = \frac{1}{\gamma}\frac{f(\theta_1)(1+\lambda)}{H'(\theta_1-e(\theta_1))}\left[\gamma H(\theta_1)+(1-\gamma)\bar{b}-\underline{b}\right]+(\alpha-1-\lambda)F(\theta_1), \tag{A.18}$$

$$\frac{\partial s_0\left(\underline{b},\bar{b}\right)}{\partial \underline{b}} = \left(1-\frac{\alpha}{1+\lambda}\right)\frac{1}{\gamma}\frac{f(\theta_1)}{H'(\theta_1-e(\theta_1))}, \tag{A.19}$$

$$\begin{aligned}\frac{\partial \pi_{int}\left(\underline{b},\bar{b}\right)}{\partial \bar{b}} &= \frac{\gamma-1}{\gamma}\frac{f(\theta_1)}{H'(\theta_1-e(\theta_1))}(1+\lambda)[\gamma H(\theta_1)+(1-\gamma)\bar{b}-\underline{b}]\\ &+ \frac{(1+\lambda)(1-\gamma)f(\theta_2)}{H'(\theta_2-e(\theta_2))}[H(\theta_2)-\bar{b}]+(1-\gamma)(\alpha-1-\lambda)(F(\theta_2)-F(\theta_1)),\end{aligned} \tag{A.20}$$

$$\begin{aligned}\frac{\partial s_0\left(\underline{b},\bar{b}\right)}{\partial \bar{b}} &= \frac{[H'(\theta_2)-H'(\theta_2-e(\theta_2))]f(\theta_2)+f'(\theta_2)[H(\theta_2)-\bar{b}]}{[H'(\theta_2-e(\theta_2))]^2}-\frac{f(\theta_2)(1-e'(\theta_2))H''(\theta_2-e(\theta_2))}{[H(\theta_2)-\bar{b}]^{-1}[H'(\theta_2-e(\theta_2))]^3}\\ &- \left(1-\frac{\alpha}{1+\lambda}\right)\left[\frac{f(\theta_2)}{H'(\theta_2-e(\theta_2))}-\frac{\gamma-1}{\gamma}\frac{f(\theta_1)}{H'(\theta_1-e(\theta_1))}\right]\equiv m(\theta_1,\theta_2).\end{aligned} \tag{A.21}$$

Then, substituting (A.20) and (A.21) into $\partial\mathcal{L}\left(\underline{b},\bar{b}\right)/\partial\bar{b}=0$ and rearranging terms for (A.11) and (A.10) yield (A.13). Similarly, substituting (A.18) and (A.19) into $\partial\mathcal{L}\left(\underline{b},\bar{b}\right)/\partial\underline{b}=0$ and rearranging terms for (A.11) and (A.10) yield (A.14).

Once (A.13) and (A.14) are established, combining (A.12) with (A.13) implies

$$\frac{f(\theta_1)s_1(\theta_1)}{H'\left(\theta_1-e\left(\theta_1\right)\right)}=\frac{\vartheta\cdot s(\theta_1,\theta_2)}{(1+\lambda)(1-\gamma)}. \tag{A.22}$$

Then, combining (A.14) with (A.22) implies an additional necessary condition for equilibrium:

$$\left(\frac{1}{\gamma}-1\right)\frac{f(\theta_1)}{H'\left(\theta_1-e\left(\theta_1\right)\right)}=\left[\frac{F(\theta_1)H'\left(\theta_1-e\left(\theta_1\right)\right)}{f(\theta_1)s_1(\theta_1)}-\left(1-\frac{\alpha}{1+\lambda}\right)^{-1}\right]s(\theta_1,\theta_2). \tag{A.23}$$

This completes the proof of Proposition 2.2. ■

Proof of Theorem 1. If $\mathbb{H}_0$ is true, then $C_t\left(\tau;\varpi_1\right)-C_t\left(\tau p;\varpi_2\right)\equiv$ a constant for $\tau\in[0,1]$ almost sure. Consequently, $C_t\left(\tau;\varpi_1\right)-C_t\left(\tau p;\varpi_2\right)=C_t\left(\nu;\varpi_1\right)-C_t\left(\nu p;\varpi_2\right)$ almost sure. And it immediately follows from Lemma A.3 that, under $\mathbb{H}_0$

$$\begin{aligned}&\sqrt{n_f}\left\{\left[\widehat{C}\left(\tau;\varpi_1\right)-\widehat{C}\left(\tau p;\varpi_2\right)\right]-\left[\widehat{C}\left(\nu;\varpi_1\right)-\widehat{C}\left(\nu p;\varpi_2\right)\right]\right\}\\ =\;&\sqrt{n_f}\left\{\left[\widehat{C}_t\left(\tau;\varpi_1\right)-\widehat{C}_t\left(\tau p;\varpi_2\right)\right]-\left[\widehat{C}_t\left(\nu;\varpi_1\right)-\widehat{C}_t\left(\nu p;\varpi_2\right)\right]\right\}\\ &-\sqrt{n_f}\left\{\left[C_t\left(\tau;\varpi_1\right)-C_t\left(\tau p;\varpi_2\right)\right]-\left[C_t\left(\nu;\varpi_1\right)-C_t\left(\nu p;\varpi_2\right)\right]\right\}\\ \xrightarrow{\mathcal{L}}\;&(1,0)\,z\left(\tau\right)/J\left(\tau\right)-(1,1)\,z\left(\tau p\right)/J\left(\tau p\right)-(1,0)\,z\left(\nu\right)/J\left(\nu\right)+(1,1)\,z\left(\nu p\right)/J\left(\nu p\right).\end{aligned} \tag{A.24}$$

According to Theorem 1.11.1 (Extended continuous mapping) in van der Vaart and Wellner

(1996), it follows from (A.24) that, under $\mathbb{H}_0$

$$\begin{aligned} S_n &\equiv \int_0^1 n_f \left\{ \left[\widehat{C}_t(\tau;\varpi_1) - \widehat{C}_t(\tau p;\varpi_2) \right] - \left[\widehat{C}_t(\nu;\varpi_1) - \widehat{C}_t(\nu p;\varpi_2) \right] \right\}^2 d\tau \\ &\xrightarrow{\mathcal{L}} \int_0^1 \left[(1,0)\, z(\tau)/J(\tau) - (1,1)\, z(\tau p)/J(\tau p) - (1,0)\, z(\nu)/J(\nu) + (1,1)\, z(\nu p)/J(\nu p) \right]^2 d\tau. \end{aligned}$$

Note that the asymptotic distribution above is unaffected when p is replaced by a $\sqrt{n}$-consistent estimator $\hat{p}$, which verifies Part (i) of Theorem 1.

To verify Part (ii) of Theorem 1, define $\Delta(\tau) \equiv [C_t(\tau;\varpi_1) - C_t(\tau p;\varpi_2)] - [C_t(\nu;\varpi_1) - C_t(\nu p;\varpi_2)]$. It follows from the support for $C|W$ being bounded that $\Delta(\cdot)$ is bounded on $[0,1]$. Moreover, under any fixed alternative, $|\Delta(\cdot)|$ is bounded away from zero on a set with nonzero measure. Consequently, we have

$$0 < \int_0^1 [\Delta(\tau)]^2 \, d\tau < +\infty. \tag{A.25}$$

It follows from (A.25), the weak law of large number and the stochastic equicontinuity implied by Lemma A.3 that

$$\left[\widehat{C}_t(\tau;\varpi_1) - \widehat{C}_t(\tau p;\varpi_2) \right] - \left[\widehat{C}_t(\nu;\varpi_1) - \widehat{C}_t(\nu p;\varpi_2) \right] \xrightarrow{p} \Delta(\tau)$$

uniformly in $\tau \in [0,1]$, which, by Theorem 1.11.1 in van der Vaart and Wellner (1996), implies

$$\frac{1}{n_f} S_n \xrightarrow{p} \int_0^1 [\Delta(\tau)]^2 \, d\tau.$$

Note that the convergence in probability result above is unaffected when p in S_n is replaced by a $\sqrt{n}$-consistent estimator $\hat{p}$, in which case S_n is changed to T_n. I.e., $\frac{1}{n_f} T_n \xrightarrow{p} \int_0^1 [\Delta(\tau)]^2 \, d\tau$, which, together with (A.25), show that T_n diverges to ∞ in probability at a rate of $O_p(n_f)$. This completes the proof. ■

Proof of Lemma 4.1. Denote by $(\widetilde{C}, \widetilde{D}^{FF}, \widetilde{D}^{CF}, \widetilde{\overline{B}}, \widetilde{\underline{B}})$ the underlying variables that correspond to $\widetilde{\mathcal{S}}$. The equivalence between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ can be obtained by taking a linear transformation that $\widetilde{\theta} = \xi_1 \theta$ with $\xi_1 > 0$. To do this, let us first consider a general linear transformation that $\widetilde{\theta} = \xi_0 + \xi_1 \theta$ with $(\xi_0, \xi_1) \in \mathbb{R}_+^2$, then the distribution of $\widetilde{\theta}$ is $\widetilde{F}(\cdot) = F((\cdot - \xi_0)/\xi_1)$. To justify the observational equivalence, we need to show that $(D^{FF}, D^{CF}, C, \overline{B}, \underline{B}) = (\widetilde{D}^{FF}, \widetilde{D}^{CF}, \widetilde{C}, \widetilde{\overline{B}}, \widetilde{\underline{B}})$, and that the equality (2.16) holds under the structure $\widetilde{\mathcal{S}}$. Let $\widetilde{\theta}_l = \xi_0 + \xi_1 \theta_l$ and $\widetilde{\theta}_u = \xi_0 + \xi_1 \theta_u$, then for any $\widetilde{\theta}$, $\widetilde{\theta} \leq \widetilde{\theta}_l$ is equivalent to $\theta \leq \theta_l$, which implies that $\widetilde{D}^{FF} = D^{FF}$. Similarly, we have $\widetilde{D}^{CF} = D^{CF}$. Note that

$$\widetilde{\psi}'(\widetilde{e}^*) = \widetilde{H}'(\widetilde{\theta} - \widetilde{e}^*) \Rightarrow \psi'[(\widetilde{e}^* - \xi_0)/\xi_1] = H'[(\widetilde{\theta} - \widetilde{e}^*)/\xi_1],$$

which leads to $\widetilde{e}(\widetilde{\theta}) = \xi_0 + \xi_1 e(\theta)$. For those with FP contracts,

$$\begin{aligned}
\widetilde{C} &= \widetilde{H}(\widetilde{\theta} - \widetilde{e}(\widetilde{\theta}^*)) = H[(\xi_1\theta - \xi_1 e(\theta^*))/\xi_1] = H(\theta - e(\theta^*)) = C, \\
\widetilde{\overline{B}} &= \widetilde{H}(\widetilde{\theta}_u - \widetilde{e}(\widetilde{\theta}_u)) + \widetilde{\psi}(\widetilde{e}(\widetilde{\theta}_u)) = H((\xi_1\theta_u - \xi_1 e(\theta_u))/\xi_1) + \psi(\xi_1 e(\theta^*)/\xi_1) = \overline{B}, \\
\widetilde{\underline{B}} &= r[\widetilde{H}(\widetilde{\theta}_l - \widetilde{e}(\widetilde{\theta}_l)) + \widetilde{\psi}(\widetilde{e}(\widetilde{\theta}_l))] + (1-r)\overline{B} \\
&= r[H(\theta_l - e(\theta_l)) + \psi(e(\theta_l))] + (1-r)\overline{B} = \underline{B}.
\end{aligned}$$

For those associated with CR contracts, since $\widetilde{C} = \widetilde{H}(\widetilde{\theta}) = H(\widetilde{\theta}/\xi_1) = H((\xi_0 + \xi_1\theta)/\xi_1)$, then $\widetilde{C} = C$ is equivalent to $\xi_0 = 0$ by noting that $C = H(\theta)$. In what follows, we just need to consider that $\widetilde{\theta} = \xi_1\theta$. Since $\widetilde{f}(\widetilde{\theta}_t^*) = \partial\widetilde{F}(\widetilde{\theta}_t^*)/\partial\widetilde{\theta}_t^* = \partial F(\widetilde{\theta}_t^*/\xi_1)/\partial\widetilde{\theta}_t^* = f(\widetilde{\theta}_t^*/\xi_1)/\xi_1 = f(\theta_t^*)/\xi_1$ (for $t = 1, 2$), we have

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{\widetilde{F}(\widetilde{\theta}_u) - \widetilde{F}(\widetilde{\theta}_l)}{\widetilde{f}(\widetilde{\theta}_u)} = \xi_1\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{F(\theta_u) - F(\theta_l)}{f(\theta_u)} = \xi_1\frac{H(\theta_u) - \overline{B}}{H'(\theta_u - e(\theta_u, w))}$$

$$\frac{\widetilde{H}(\widetilde{\theta}_u) - \widetilde{\overline{B}}}{\widetilde{H}'(\widetilde{\theta}_u - \widetilde{e}(\widetilde{\theta}_u))} = \frac{\widetilde{H}(\widetilde{\theta}_u) - \widetilde{\overline{B}}}{\widetilde{\psi}'(\widetilde{e}(\widetilde{\theta}))} = \frac{H(\theta_u) - \overline{B}}{\psi'(e^*)/\xi_1} = \xi_1\frac{H(\theta_u) - \overline{B}}{H'(\theta_u - e(\theta_u))}.$$

The two equations above immediately imply that,

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\frac{\widetilde{F}(\widetilde{\theta}_u) - \widetilde{F}(\widetilde{\theta}_l)}{\widetilde{f}(\widetilde{\theta}_u)} = \frac{\widetilde{H}(\widetilde{\theta}_u) - \widetilde{\overline{B}}}{\widetilde{H}'(\widetilde{\theta}_u - \widetilde{e}(\widetilde{\theta}_u))}.$$

This completes the proof. ■

Proof of Proposition 4.1. In this proof, we discuss in details the assumptions required to identify the distribution of E using Schennach and Hu (2013), as well as other related issues.

In addition to Assumption 4.1, for $t = 1$ or 2 (fixed), we impose a few mild regularity conditions to achieve identification, as follows:
(a) The characteristic function of $V_{t,1}$ and $V_{t,2}$ do not vanish anywhere.
(b) The distribution of E_t admits a uniformly bounded density $f_{E_t}(\cdot)$ with respect to the Lebesgue measure that is supported on an interval (which may be infinite).
(c) The function $m_t(\cdot)$ is continuously differentiable over the interior of the support of E_t.
(d) The set $\chi \equiv \{e : m_t(e) = 0\}$ has at most a finite number of elements $e_1, \cdots, e_s$ If χ is nonempty, $f_{E_t}(\cdot)$ is continuous and non-vanishing in a neighborhood of each e_l, $l = 1, \cdots s$.

Part (a) is a widely used assumption in the literature of measurement errors. Most of the commonly encountered distributions satisfy this condition, with the notable exceptions being the uniform and the triangular distributions. Parts (b)-(c) are standard smoothness constraints. Part (d) states that we allows for non-monotone function $m(\cdot)$, but rules out functions that are constant over an interval (not reduced to a point) or that exhibit an infinite number of oscillations. Nevertheless, this condition is sufficiently flexible to encompass most specifications of practical interest.

Given Assumption 4.1 and the conditions (a)-(d), we have the following results on the identification of $f_E(\cdot)$.

1. If $m_t(\cdot)$ is not of the form $m_t(e) = a + b\ln(\exp(ce) + d)$ for some constants $a, b, c, d \in \mathbb{R}$. Then, $f_{E_t}(\cdot)$ and $m_t(\cdot)$ are nonparametrically identified.

2. If $m_t(\cdot)$ is linear, i.e., of the form above with $d = 0$, $f_{E_t}(\cdot)$ and $m_t(\cdot)$ are identified.

Note that if $m_t(\cdot)$ is linear, Schennach and Hu (2013) show that neither $f_{E_t}(\cdot)$ nor $m_t(\cdot)$ is identified if and only if E_t is normally distributed and either $V_{t,1}$ or $V_{t,2}$ can be decomposed as a summation of two variables with one of them being normally distributed. However, in our setting, the effort e is assumed to be non-negative, so it cannot be normally distributed. Thus both $f_{E_t}(\cdot)$ and $m_t(\cdot)$ are identified. According Theorem 1 in Schennach and Hu (2013), the only scenario where we cannot identify $f_{E_t}(\cdot)$ and $m_t(\cdot)$ is where: (i) $m_t(\cdot)$ is of the form $m_t(e) = a + b\ln(\exp(ce) + d)$ with $d \neq 0$; (ii) E_t has a density of the form $f_{E_t}(e) = A\exp(-B\exp(Ce) + CDe)(\exp(De) + G)^{-W}$ with $C \in \mathbb{R}$, $A, B, D, G, W \in [0, \infty)$; (iii) $V_{t,2}$ can be written as a summation of two random variables with one of them being a Type I extreme value variable. ■

Proof of Proposition 4.2. According to Corollary 3.1, under FP, the range of $C_t|W = \varpi_j$, i.e., the period-t cost conditioned on $W = \varpi_j$, for $j = 1, 2$, is as follows:

$$C_t|W = \varpi_j \in [C_t(0; \varpi_j), C_t(1; \varpi_j)] = \left[H(\underline{\theta} - e(\underline{\theta}; \varpi_j)), H\left(\theta_t^{1*} - e\left(\theta_t^{1*}; \varpi_j\right)\right)\right].$$

Note that it follows from Assumption 4.2 (ii) and the f.o.c. (2.2) that $e(\underline{\theta}; \varpi_1) = e(\underline{\theta}; \varpi_2)$. Consequently, it holds that $C_t(0; \varpi_1) = C_t(0; \varpi_2) = \underline{C}$.

Next, we verify following claim: For any given $\tau \in (0, p_{t,1}]$ it holds that,

$$e(\theta(\tau); \varpi_1) < e(\theta(\tau); \varpi_2) \quad \text{or, equivalently,} \quad E_t(\tau/p_{t,1}; \varpi_1) < E_t(\tau/p_{t,2}; \varpi); \tag{A.26}$$

$$c(\theta(\tau); \varpi_1) > c(\theta(\tau); \varpi_2) \quad \text{or, equivalently,} \quad C_t(\tau/p_{t,1}; \varpi_1) > C_t(\tau/p_{t,2}; \varpi_2). \tag{A.27}$$

(A.26) can be verified by contradiction, as follows: Suppose $e(\theta(\tau); \varpi_1) \geq e(\theta(\tau); \varpi_2)$, which immediately implies that $\theta(\tau) - e(\theta(\tau); \varpi_1) \leq \theta(\tau) - e(\theta(\tau); \varpi_2)$. It then follows from $H''(\cdot) > 0$ (by Assumption 3.2 (i)) that

$$H'(\theta(\tau) - e(\theta(\tau); \varpi_1)) \leq H'(\theta(\tau) - e(\theta(\tau); \varpi_2)),$$

which, together with (2.2) (i.e., $H'(\theta - e(\theta)) = \psi'(e(\theta))$), imply that

$$\psi_1'(e(\theta(\tau); \varpi_1)) \leq \psi_2'(e(\theta(\tau); \varpi_2)) < \psi_1'(e(\theta(\tau); \varpi_2)), \tag{A.28}$$

where the strict inequality follows from Assumption 4.2 (ii) and the fact that $e(\theta(\tau); \varpi_2) > e(\underline{\theta}; \varpi_2) = \underline{e}$ for any $\tau > 0$. However, since $\psi_1''(\cdot) > 0$ (by Assumption 3.2 (ii)), it is impossibly for both $e(\theta(\tau);\varpi_1) \geq e(\theta(\tau);\varpi_2)$ and $\psi_1^{'}(e(\theta(\tau);\varpi_1)) < \psi_1^{'}(e(\theta(\tau);\varpi_2))$ (as immediately implied by (A.28)) to hold. Thus, we reach a contradiction, and it has to be the case that $e(\theta(\tau);\varpi_1) < e(\theta(\tau);\varpi_2)$, as claimed by (A.26). Given that (A.26) has just been verified, (A.27) follows immediate from (A.26) and the fact that $c(\theta(\tau);\varpi_j) = H(\theta(\tau) - e(\theta(\tau);\varpi_j))$ for all $\tau \leq p_{t,1}(= \min_{j=1,2} p_{t,j})$.

Now equipped with (A.26) and (A.27), we proceed with the proof. Note that, for any $c \in [C_t(0;\varpi_1), C_t(1;\varpi_1)]$, there exist a unique $\tau_0(c) \in [0,1]$ s.t. $c = C_t(t_0(c);\varpi_1)$. Moreover, $\tau_0(c) = 0$ for $c = \underline{C} = C_t(0;\varpi_1)$, and $\tau_0(c) \in (0,1]$ for $c \in (C_t(0;\varpi_1), C_t(1;\varpi_1)]$.

Consequently, for a given $c \in (C_t(0;\varpi_1), C_t(1;\varpi_1)]$, it follows from (A.26) that

$$c = C_t(\tau_0(c);\varpi_1) > C_t(\tau_0(c) \cdot p_{t,1}/p_{t,2};\varpi_2).$$

Similarly, there exists a unique $\tau_1(c) \in (0, \tau_0(c))$ s.t.

$$C_t(\tau_1(c);\varpi_1) = C_t(\tau_0(c).p_{t,1}/p_{t,2};\varpi_2) > C_t(\tau_1(c).p_{t,1}/p_{t,2};\varpi_2).$$

Iteratively, given $\tau_k(c) \in (0, \tau_{k-1}(c))$, there exists a unique $\tau_{k+1}(c) \in (0, \tau_k(c))$ s.t.

$$C_t(\tau_{k+1}(c);\varpi_1) = C_t(\tau_k(c).p_{t,1}/p_{t,2};\varpi_2) > C_t(\tau_{k+1}(c).p_{t,1}/p_{t,2};\varpi_2). \tag{A.29}$$

And we end up obtaining a strictly decreasing sequence $\{\tau_k(c)\}_k \subseteq (0,1]$. The boundedness and monotonicity of $\{\tau_k(c)\}_k$ imply that a limit for the sequence exits as $k \to \infty$, denoted by

$$\underline{\tau}(c) \equiv \lim_{k\to\infty} \tau_k(c).$$

By the continuity of $C_t(\cdot;\varpi_j)$ for both $j = 1,2$ (implied by the conditional distribution of C_t on W being continuous), taking limit on both side of the equality from (A.29), i.e., $C_t(\tau_{k+1}(c);\varpi_1) = C_t(\tau_k(c).p_{t,1}/p_{t,2};\varpi_2)$, yields

$$C_t(\underline{\tau}(c);\varpi_1) = C_t(\underline{\tau}(c).p_{t,1}/p_{t,2};\varpi_2), \tag{A.30}$$

which is true only if $\underline{\tau}(c) = 0$. Therefore, we have just shwon that $\tau_k(c) \to 0$ as $k \to \infty$, for any given $c \in (C_t(0;\varpi_1), C_t(1;\varpi_1)]$.

To summarize, for any given $c \in [C_t(0;\varpi_1), C_t(1;\varpi_1)]$:
(i) If $c = C_t(0;\varpi_1)$, then we have $H^{-1}(c) = H^{-1}(\underline{C}) = \underline{\theta} - \underline{e}$;
(ii) If $c \in (C_t(0;\varpi_1), C_t(1;\varpi_1)]$, then we have

$$\begin{aligned}
H^{-1}(c) &= H^{-1}(C_t(\tau_0(c);\varpi_1)) \\
&= H^{-1}(C_t(\tau_0(c) \cdot p_{t,1}/p_{t,2};\varpi_2)) + \Delta\tilde{E}_t(\tau_0(c) \cdot p_{t,1}) \\
&= H^{-1}(C_t(\tau_1(c);\varpi_1)) + \Delta\tilde{E}_t(\tau_0(c) \cdot p_{t,1}) \\
&= H^{-1}(C_t(\tau_1(c) \cdot p_{t,1}/p_{t,2};\varpi_2)) + \Delta\tilde{E}_t(\tau_1(c) \cdot p_{t,1}) + \Delta\tilde{E}_t(\tau_0(c) \cdot p_{t,1}) \\
&= \cdots \\
&= H^{-1}(C_t(\tau_m(t) \cdot p_{t,1}/p_{t,2};\varpi_2)) + \sum_{k=0}^{m} \Delta\tilde{E}_t(\tau_k(c) \cdot p_{t,1}).
\end{aligned}$$

Rearranging terms for the equation above yields

$$\sum_{k=0}^{m} \Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right) \quad = \quad H^{-1}\left(c\right) - H^{-1}\left(C_t\left(\tau_m(c)\cdot p_{t,1}/p_{t,2};\varpi_2\right)\right). \tag{A.31}$$

Letting $m \to \infty$ on both side of (A.31) and rearranging terms yields

$$\begin{aligned} H^{-1}(c) &= H^{-1}\left(C_t\left(\underline{\tau}\left(c\right)\cdot p_{t,1}/p_{t,2};\varpi_2\right)\right) + \sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right) \\ &= H^{-1}\left(C_t\left(0;\varpi_2\right)\right) + \sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right) \\ &= \underline{\theta} - \underline{e} + \sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right), \end{aligned}$$

where the second equation follows from $\underline{\tau}\left(c\right) = \lim_{k\to\infty}\tau_k(c) = 0$, which we have already shown. This completes the proof. ■

Proof of Proposition 5.1. The proposed $\widehat{F}(\cdot)$, $\widehat{\psi}_1(\cdot)$, $\widehat{\psi}_2(\cdot)$, $\widehat{\alpha/(1+\lambda)}$ and $\hat{\gamma}$ are all plug-in type estimators based on the estimated inverse cost function $\widehat{H}^{-1}(\cdot)$. We focus to verify the consistency of $\widehat{H}^{-1}(\cdot)$, which is the key to guaranteeing consistency of all other estimators.

Recall that the sample criterion function for estimating $H^{-1}(\cdot)$ is

$$Q_n\left(H^{-1}\right) = \int_0^1 \left[H^{-1}\left(\widehat{C}_2(\tau\cdot\hat{p}_{2,1}/\hat{p}_{2,2};\varpi_2)\right) - H^{-1}\left(\widehat{C}_2\left(\tau;\varpi_1\right)\right) + \Delta\widehat{\tilde{E}_2}(\tau\cdot\hat{p}_{2,1})\right]^2 d\tau.$$

Define the population criterion function as

$$Q\left(H^{-1}\right) = \int_0^1 \left[H^{-1}\left(C_2(\tau\cdot p_{2,1}/p_{2,2};\varpi_2)\right) - H^{-1}\left(C_2\left(\tau;\varpi_1\right)\right) + \Delta\tilde{E}_2(\tau\cdot p_{2,1})\right]^2 d\tau,$$

Also recall that the estimated inverse cost function $\widehat{H}^{-1}$ is characterized by

$$\widehat{H}^{-1} = \underset{H^{-1}\in\mathcal{H}_n:\ H^{-1}(\underline{c})=\underline{\theta}-\underline{e}}{\operatorname{argmin}} Q\left(H^{-1}\right).$$

And it is easy to see that the true function H_0^{-1} can be equivalently characterized by

$$H_0^{-1} = \underset{H^{-1}\in\mathcal{H}:\ H^{-1}(\underline{c})=\underline{\theta}-\underline{e}}{\operatorname{argmin}} Q\left(H^{-1}\right),$$

where $\mathcal{H}$ is the (infinite-dimensional) parameter space for H^{-1}. Since we only consider smooth functions for H^{-1}, we specify $\mathcal{H}$ to be a bounded subset of the following Sobolev space:

$$\mathcal{W}^s \equiv \left\{H^{-1}:[\underline{c},\bar{c}]\to\mathbb{R} \mid\ H^{-1} \text{ is 3-times differentiable and } \left\|H^{-1}\right\|_s \leq \infty\right\}$$

with $\|\cdot\|_s$ being a commonly used norm for Sobolev spaces, defined as

$$\left\|H^{-1}\right\|_s^2 \equiv \sum_{\epsilon=1}^{3} \int_{\underline{c}}^{\bar{c}} \left|\partial^{\epsilon} H^{-1}(c)/\partial x_j^{\epsilon}\right|^2 dc.$$

Specifically, $\mathcal{H} \equiv \left\{H^{-1} \in \mathcal{W}^s : \left\|H^{-1}\right\|_s \leq A\right\}$ for some constant $A < \infty$. It can be verified by following the arguments of Freyberger and Masten (2019) that $\mathcal{H}$ is compact under the L^2-norm.

Note that, under point-identification, H^{-1} is the unique minimizer of $Q(\cdot)$ over the functional space $\mathcal{H}$, and $Q(\beta_0) = 0$. In addition, $Q\left(H^{-1}\right)$ is continuous under the L^2-norm by construction. Consequently, for any given $\varepsilon > 0$, it follows from the compactness of $\mathcal{H}$ and the continuity of $Q(\cdot)$ under the L^2-norm that

$$\inf_{H^{-1} \in \mathcal{H}: \left\|H^{-1} - H_0^{-1}\right\|_{L^2} \geq \varepsilon} Q\left(H^{-1}\right) > 0 = Q\left(H_0^{-1}\right). \tag{A.32}$$

Also note that $\{\hat{p}_{2,1}, \hat{p}_{2,2}\}$ are estimators directly based on corresponding empirical distributions, and that $\left\{\widehat{C}_2(\cdot;\varpi_1), \widehat{C}_2(\cdot;\varpi_2), \widehat{E}_2(\cdot;\varpi_1), \widehat{E}_2(\cdot;\varpi_2)\right\}$ are obtained from standard quantile regression. Their consistency are guarantee by the random sample condition and a few regularity conditions (for the quantile regression) which are satisfied in our econometric setting. Consequently, the compactness of $\mathcal{H}$ and the continuity of $Q(\cdot)$ under the L^2-norm, and the consistency of $\left\{\hat{p}_{2,1}, \hat{p}_{2,2}, \widehat{C_2}(\tau;\varpi_1),\ \widehat{C_2}(\tau \cdot p_{2,1}/p_{2,2};\varpi_2), \widehat{E}_2(\tau;\varpi_1),\ \widehat{E}_2(\tau \cdot p_{2,1}/p_{2,2};\varpi_2)\right\}$ for all $\tau \in [0,1]$ guarantee that

$$\sup_{H^{-1} \in \mathcal{H}} |Q_n(H^{-1}) - Q(H^{-1})| \xrightarrow{p} 0. \tag{A.33}$$

(A.32) and (A.33) together imply that $\left\|\widehat{H}^{-1} - H_0^{-1}\right\|_{L^2} \xrightarrow{p} 0$ according to 5.7 Theorem in van der Vaart (1999). ■

Online Supplement for
"A Structural Analysis of Simple Contracts"
(NOT FOR PUBLICATION)

Yonghong An[a], Shengjie Hong[b], Daiqiang Zhang[c]

[a] Department of Economics, Texas A&M University

[b] School of Economics, Central University of Finance and Economics

[c] Department of Economics, University at Albany

This online supplement consists of Appendices B, C and D. In Appendix B, we provide additional details on the empirical study. In Appendix C, we provide proofs of technical lemmas A.1 – A.3 listed in Appendix A. Appendix D collects additional supplementary contents, including an alternative identification result for $H^{-1}(\cdot)$ (briefly mentioned in Section 4.2), and the LCC test and identification results for the single-period setting.

B Additional details on the empirical study

We first discuss identification of model primitives under the LCC, then we present details on the estimation of welfare.

B.1 Identification under the LCC

Here we discuss how the identification strategies can be simplified under the LCC, for the most part. Though as we discuss below, under the LCC, more specific assumptions on the disutility functions are needed, due to a lack of variation in the optimal effort. Like in the main context of the paper, our discussion here is focused on the single-period setting.

As discussed in Section 3.2, a linear cost function $H(\theta - e) = \beta(\theta - e)$ induces a constant optimal effort under FP, regardless of θ. Thus, the two effort proxies X and Y (for identifying/estimating the distribution of effort) are no longer needed, as the distribution now degenerates to a single point (with probability mass one) e^{*j} conditioned on $W = \varpi_j$, for $j = 1, 2$. Note that e^{j*} is characterized by the f.o.c. $\psi_j'\left(e^{j*}\right) = \beta$. And the cutoff type θ^{j*} now satisfies

$$b^{j*} = \beta\left(\theta^{j*} - e^{j*}\right) + \psi_j\left(e^{j*}\right) \tag{B.1}$$

Still conditioned on $W = \varpi_j$, denote by $\bar{c}^j$ and $\underline{c}_{CR}^j$ the cost upper bound under FP and the cost lower bound under CR, respectively. We have

$$\bar{c}^j = \beta\left(\theta^{j*} - e^{j*}\right) \quad \text{and} \quad \underline{c}_{CR}^j = \beta\theta^{j*}. \tag{B.2}$$

Note that the lack of variation in optimal effort makes it impossible to identify $\left\{e^{j*}, \psi_1(\cdot), \psi_2(\cdot)\right\}$ without additional assumptions. For this reason, we normalize $\psi_1(e) = e^2$, and adopt a simple parameterization $\psi_2(e) = \kappa e^2$ for ψ_2, which satisfies Assumption 3.2, and is widely used in the literature. See, for example, Laffont and Tirole (1988), Rogerson (2003), Chu and Sappington (2007), and Battaglini (2007). These specifications on $\{\psi_1(\cdot), \psi_2(\cdot)\}$, together with (B.1) and (B.2), identify e^{1*} as

$$e^{1*} = \left(b^{1*} - \bar{c}^1\right)^{1/2}. \tag{B.3}$$

Consequently, accoding to (B.2) and (B.3), β is identified as

$$\beta = \frac{\underline{c}^1_{CR} - \bar{c}^1}{e^{1*}} = \frac{\underline{c}^1_{CR} - \bar{c}^1}{\left(b^{1*} - \bar{c}^1\right)^{1/2}}.$$

With β being identified, θ^{j*}, for $j = 1, 2$, are immediately identified from the second equation in (B.2) as $\theta^{j*} = \underline{c}^j_{CR}/\beta$. Consequently, $\left\{e^{2*}, \kappa\right\}$ are identified as the solution to

$$\begin{cases} \psi_2'\left(e^{2*}\right) = 2\kappa e^{2*} = \beta; \\ b^{2*} = \bar{c}^2 + \kappa\left(e^{2*}\right)^2. \end{cases}$$

In close forms, we have

$$\begin{aligned} e^{2*} &= \frac{2\left(b^{2*} - \bar{c}^2\right)}{\beta} = \frac{2\left(b^{2*} - \bar{c}^2\right)\left(b^{1*} - \bar{c}^1\right)^{1/2}}{\underline{c}^1_{CR} - \bar{c}^1}, \\ \kappa &= \frac{\beta}{2e^{2*}} = \frac{\left(\underline{c}^1_{CR} - \bar{c}^1\right)^2}{4\left(b^{1*} - \bar{c}^1\right)\left(b^{2*} - \bar{c}^2\right)}. \end{aligned}$$

With β and e^{j*} for $j = 1, 2$ being identified, we can recover the underlying θ associated with any observed cost c conditioned on $W = \varpi_j$ as

$$\theta = \begin{cases} c/\beta + e^{j*}, \text{ under FP}; \\ c/\beta, \text{ under CR}, \end{cases}$$

based on which the distribution of θ is identified. The identifications of α remains unchanged.

We note the followings at the end of this discussion: (i) The identification steps above are easily implementable for conducting estimation under the LCC; (ii)Following the existing literature, we specify $\psi_2(\cdot)$ as $\psi_2(e) = \kappa e^2$ under the LCC. As explained, this is due to the lack of variation in the optimal effort under the LCC. The necessity of adopting such a simplifying specification is further illustrated by the fact that identification would fail under a slightly more general specification $\psi_2(e) = \kappa_1 e + \kappa_2 e^2$. Specifically, under this alternative specification, $\left(e^{2*}, \kappa_1, \kappa_2\right)$ are not jointly identifiable. This is because these three parameters are characterized by merely two equations, which are $\beta = \kappa_1 + 2\kappa_2 e^{2*}$ (implied by the f.o.c. regarding the e^{2*}) and $b^{2*} = \bar{c}^2 + \kappa_1 e^{2*} + \kappa_2 \left(e^{2*}\right)^2$.

The identification results above are constructive, and we follow the identification procedure to estimate the model primitives. ■

B.2 Estimation of welfare

For $z = 1, 2$ and $j = 1, 2$, let $SW(z)$ denote the welfare conditional on $Z = z$, and $SW^j(z)$ denote the welfare conditional on $Z = z$ and $W = \varpi_j$. By (2.5), we have

$$\begin{aligned} SW^j(z) &= S - (1+\lambda) Q^j(z) + \alpha U^j(z), \\ SW(z) &= \sum_{j=1,2} SW^j(z) \cdot \Pr(W = \varpi_j) \equiv S - (1+\lambda) Q(z) + \alpha U(z), \end{aligned}$$

with $Q(z) \equiv \sum_{j=1,2} Q^j(z) \cdot \Pr(W = \varpi_j)$ and $U(z) \equiv \sum_{j=1,2} U^j(z) \cdot \Pr(W = \varpi_j)$. For a given $Z = z$, $Q^j(z)$ and $U^j(z)$ for $j = 1, 2$ are estimated as follows.

$$\begin{aligned} Q^j(z) &\equiv \underline{b}^j F_{\theta|z}\left(\theta_1^{j*}(z)|z\right) + \int_{\theta_1^{j*}(z)}^{\theta_2^{j*}(z)} \left[rH(\theta) + (1-r)\bar{b}^j\right] dF_{\theta|z}(\theta|z) + \int_{\theta_2^{j*}(z)}^{\bar{\theta}(z)} H(\theta)\, dF_{\theta|z}(\theta|z), \\ U^j(z) &\equiv \int_{\underline{\theta}(z)}^{\theta_1^{j*}(z)} \left[\underline{b}^j - H(\theta - e(\theta)) - \psi(e(\theta))\right] dF_{\theta|z}(\theta|z) \\ &\quad + (1-r)\int_{\theta_1^{j*}(z)}^{\theta_2^{j*}(z)} \left[\bar{b}^j - H(\theta - e(\theta)) - \psi(e(\theta))\right] dF_{\theta|z}(\theta|z). \end{aligned}$$

$\theta_1^{j*}(z)$ and $\theta_2^{j*}(z)$ represent the two cutoff values for the innate cost conditioned on $Z = z$ and $W = \varpi_j$, for $z = 1, 2$ and $j = 1, 2$. Specifically, for given $\{z, j\}$, $\theta_1^{j*}(z)$ and $\theta_2^{j*}(z)$ are computed as the numerically solution to equations (3.4) and (3.5) conditioned on $Z = z$ and $W = \varpi_j$. Recall that

$$SW \equiv S - (1+\lambda) Q + \alpha U = \sum\nolimits_{z=1}^{2} SW(z) \Pr(Z = z). \tag{B.4}$$

Accordingly, we estimate SW up to a constant S, by

$$\widehat{SW} = S - \left(1 + \hat{\lambda}\right) \sum\nolimits_{z=1}^{2} \widehat{Q}(z) \Pr(Z = z) + \hat{\alpha} \sum\nolimits_{z=1}^{2} \widehat{U}(z) \Pr(Z = z),$$

with $\widehat{Q}(z) = \sum_{j=1,2} \widehat{Q}^j(z) \cdot \widehat{\Pr}(W = \varpi_j)$ and $\widehat{U}(z) = \sum_{j=1,2} \widehat{U}^j(z) \cdot \widehat{\Pr}(W = \varpi_j)$. The estimators $\widehat{Q}^j(z)$ and $\widehat{U}^j(z)$ are constructed by their definition above. Under the LCC, SW is estimated similarly.

C Proofs of Lemmas A.1 – A.3

Proof of Lemma A.1. The proof follows the same steps as the proof of Proposition 3 in Gagnepain et al. (2013), except for a more complex formula for calculating the expected welfare (detailed by (2.10) – (2.12 in Section 2.3). Such a difference in calculating the expected welfare is due to that a general cost function is in place of the special form $H(\theta - e) = \theta - e$ adopted by Gagnepain et al. (2013), yet is nonessential for proving Lemma A.1. Thus, the proof is skipped. ■

Proof of Lemma A.2. Consider any initial contract $\mathbf{b}^0 = \left(b_1^{FF,0}, b_2^{FF,0}, b_2^{CF,0}\right) \equiv \left(b_1^{FF,0}, R^0\right)$ and the renegotiated offer $\widetilde{R} = \left(\widetilde{b}_2^{FF}, \widetilde{b}_2^{CF}\right)$ that satisfies (A.1).By offering C^0 followed by a renegotiated $\widetilde{R}$, the principal anticipate the expected welfare for the second period to be

$$\begin{aligned} SW_2\left(\mathbf{b}^0, \widetilde{R}\right) &= \int_{\underline{\theta}}^{\theta_1(\mathbf{b}^0)} \left[S - (1+\lambda)\widetilde{b}_2^{FF} + \alpha\left(\widetilde{b}_2^{FF} - H\left(\theta - e(\theta)\right) - \psi\left(e(\theta)\right)\right)\right] dF(\theta) \\ &+ \int_{\theta_1(\mathbf{b}^0)}^{\theta_2\left(\widetilde{b}_2^{CF}\right)} \left[S - (1+\lambda)\widetilde{b}_2^{CF} + \alpha\left(\widetilde{b}_2^{CF} - H\left(\theta - e(\theta)\right) - \psi\left(e(\theta)\right)\right)\right] dF(\theta) \\ &+ \int_{\theta_2\left(\widetilde{b}_2^{CF}\right)}^{\bar{\theta}} \left[S - (1+\lambda)H(\theta)\right] dF(\theta). \end{aligned} \tag{C.1}$$

Now consider a renegotiation-proof offer $\mathbf{b} = \left(b_1^{FF}, b_2^{FF}, b_2^{CF}\right) \equiv \left(b_1^{FF}, R\right)$. By Lemma A.1, $R = \left(b_2^{FF}, b_2^{CF}\right)$ solves the following constrained maximization problem:

$$\max_{\widetilde{R}=\left(\widetilde{b}_2^{FF}, \widetilde{b}_2^{CF}\right)} SW_2\left(\mathbf{b}, \widetilde{R}\right) \quad \text{subject to (A.1).} \tag{C.2}$$

It follows from the condition $\alpha < 1 + \lambda$ that (A.1) is binding for any solution to (C.2). Consequently, the f.o.c. of the optimization problem (C.2) with respect to $\widetilde{b}_2^{CF}$ at $\widetilde{b}_2^{CF} = b_2^{CF}$ is

$$\begin{aligned} 0 &= \partial SW_2\left(\mathbf{b}, \widetilde{R}\right) / \partial \widetilde{b}_2^{CF}|_{\widetilde{b}_2^{CF} = b_2^{CF}} + \eta \\ &= \theta_2'\left(b_2^{CF}\right)\left\{S - (1+\lambda)b_2^{CF}\right. \\ &\quad \left. + \alpha\left[b_2^{CF} - H(\theta_2\left(b_2^{CF}\right) - e(\theta_2\left(b_2^{CF}\right))) - \psi(e(\theta_2\left(b_2^{CF}\right)))\right]\right\} f(\theta_2\left(b_2^{CF}\right)) \\ &\quad + \int_{\theta_1(\mathbf{b})}^{\theta_2\left(b_2^{CF}\right)} (\alpha - 1 - \lambda) f(\theta) d\theta - \theta_2'\left(b_2^{CF}\right)\left[S - (1+\lambda)H(\theta_2\left(b_2^{CF}\right))\right] f(\theta_2\left(b_2^{CF}\right)) + \eta \\ &= \theta_2'\left(b_2^{CF}\right)(1+\lambda)\left[H(\theta_2\left(b_2^{CF}\right)) - b_2^{CF}\right] f(\theta_2\left(b_2^{CF}\right)) \\ &\quad + \int_{\theta_1(\mathbf{b})}^{\theta_2\left(b_2^{CF}\right)} (\alpha - 1 - \lambda) f(\theta) d\theta + \eta, \end{aligned} \tag{C.3}$$

where $\eta > 0$ is the Lagrange multiplier corresponding to the binding constraint (A.1).

It also holds that $b_2^{CF} = H\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right) + \psi\left(e\left(\theta_2\left(b_2^{CF}\right)\right)\right)$, of which taking derivative w.r.t. b_2^{CF} on both sides and rearranging terms yield

$$\begin{aligned}\theta_2'\left(b_2^{CF}\right) &= \Big\{H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right) \\ &\quad - \Big[H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right) - \psi'\left(e\left(\theta_2\left(b_2^{CF}\right)\right)\right)\Big] e'\left(\theta_2\left(b_2^{CF}\right)\right)\Big\}^{-1} \\ &= \frac{1}{H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right)}, \end{aligned} \tag{C.4}$$

where the second equality follows from the f.o.c. regarding the optimal effort, specifically, $H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right) - \psi'\left(e\left(\theta_2\left(b_2^{CF}\right)\right)\right) = 0$.

Taking the fact that $\eta > 0$, substituting (C.4) into (C.3) and rearranging terms yield

$$\left(1 - \frac{\alpha}{1+\lambda}\right)\left[F\left(\theta_2\left(b_2^{CF}\right)\right) - F\left(\theta_1\left(\mathbf{b}\right)\right)\right] \geq \frac{\left[H\left(\theta_2\left(b_2^{CF}\right)\right) - b_2^{CF}\right] f\left(\theta_2\left(b_2^{CF}\right)\right)}{H'\left(\theta_2\left(b_2^{CF}\right) - e\left(\theta_2\left(b_2^{CF}\right)\right)\right)},$$

which completes the proof of Lemma A.2. ■

Proof of Lemma A.3. Without loss of generality, set $\varpi_1 = 0$ and $\varpi_2 = 1$. So the binary variable W itself becomes a 0 and 1 dummy variable. Consequently, $W = \mathbf{1}\left(W = \varpi_2\right)$ and $w_i = \mathbf{1}\left(w_i = \varpi_2\right)$. Define

$$\begin{aligned} Q_{n_f}\left(q_1, q_2, \tau\right) &\equiv \frac{1}{n_f} \sum_{i \in \mathcal{FP}_t} \rho_\tau\left(c_i - q_1 - q_2 \cdot w_i\right), \\ Q\left(q_1, q_2, \tau\right) &\equiv \mathbb{E}\left[\rho_\tau\left(C - q_1 - q_2 \cdot W\right) | \text{FP for period-}t\right]. \end{aligned}$$

In addition, define $q(\tau) = \left(q_1(\tau), q_2(\tau)\right)' \equiv \underset{q=(q_1,q_2)'}{\operatorname{argmin}}\ Q\left(q_1, q_2, \tau\right)$. It holds that

$$\widehat{q}(\tau) = \left(\widehat{q}_1(\tau), \widehat{q}_2(\tau)\right)' = \underset{q=(q_1,q_2)'}{\operatorname{argmin}}\ Q_{n_f}\left(q_1, q_2, \tau\right). \tag{C.5}$$

Note that all conditions required by Theorem 3 in Angrist et al. (2006) hold in our setting, with c_i in place of their Y_i, and w_i in place of their X_i. Specifically, Conditions (i), (ii) of Theorem 3 in Angrist et al. (2006) are guaranteed by Assumption 3.1 (i), (iii). And Conditions (iii), (iv) of Theorem 3 in Angrist et al. (2006) hold in our setting due to the boundedness of the support for $C|W$ and the fact that w_i is a binary variable. Moreover, since the dependent variable w_i is binary, the linear quantile regressions we consider is correctly specified. Therefore, the second result of Theorem 3 in Angrist et al. (2006), specified by their Equation (15), is applicable to our setting. Accordingly, it holds that

$$J(\tau)\sqrt{n_f}\left(\widehat{q}(\tau) - q(\tau)\right) \xrightarrow{\mathcal{L}} z(\tau) \tag{C.6}$$

with $J(\tau) = \mathbb{E}\left[f_{C_t|W}(q_1(\tau) + q_2(\tau) W|\varpi_2) W\right]$, where $z(\cdot)$ is a zero mean tight Gaussian process on $L^\infty[0,1]$, and the weak convergence "$\xrightarrow{\mathcal{L}}$" is uniform in the sense of Chapter 1.3 in van der Vaart and Wellner (1996).

Consequently, it follows from Theorem 1.11.1 (Extended continuous mapping) in van der Vaart and Wellner (1996) that

$$\begin{aligned}
\sqrt{n_f}\left[\widehat{C}_t(\tau;\varpi_1) - C_t(\tau;\varpi_1)\right] &= \sqrt{n_f}\left(\widehat{q}_1(\tau) - q_1(\tau)\right) \\
&= (1,0)\sqrt{n_f}\left(\widehat{q}(\tau) - q(\tau)\right) \\
&\xrightarrow{\mathcal{L}} (1,0)\, z(\tau)/J(\tau), \qquad \text{(C.7)} \\
\sqrt{n_f}\left[\widehat{C}_t(\tau p;\varpi_2) - C_t(\tau p;\varpi_2)\right] &= \sqrt{n_f}\left[\left(\widehat{q}_1(\tau p) + \widehat{q}_2(\tau p)\right) - \left(q_1(\tau p) + q_2(\tau p)\right)\right] \\
&= (1,1)\sqrt{n_f}\left(\widehat{q}(\tau p) - q(\tau p)\right) \\
&\xrightarrow{\mathcal{L}} (1,1)\, z(\tau p)/J(\tau p), \qquad \text{(C.8)}
\end{aligned}$$

for τ on $[0,1]$. Again, according to Theorem 1.11.1 in van der Vaart and Wellner (1996), it follows from (C.7) and (C.8) that

$$\begin{aligned}
&\sqrt{n_f}\left\{\left[\widehat{C}_t(\tau;\varpi_1) - \widehat{C}_t(\tau p;\varpi_2)\right] - \left[\widehat{C}_t(\nu;\varpi_1) - \widehat{C}_t(\nu p;\varpi_2)\right]\right\} \\
&- \sqrt{n_f}\left\{\left[C_t(\tau;\varpi_1) - C_t(\tau p;\varpi_2)\right] - \left[C_t(\nu;\varpi_1) - C_t(\nu p;\varpi_2)\right]\right\} \\
= \; &\sqrt{n_f}\left[\widehat{C}_t(\tau;\varpi_1) - C_t(\tau;\varpi_1)\right] - \sqrt{n_f}\left[\widehat{C}_t(\tau p;\varpi_2) - C_t(\tau p;\varpi_2)\right] \\
&- \sqrt{n_f}\left[\widehat{C}_t(\nu;\varpi_1) - C_t(\nu;\varpi_1)\right] + \sqrt{n_f}\left[\widehat{C}_t(\nu p;\varpi_2) - C_t(\nu p;\varpi_2)\right] \\
\xrightarrow{\mathcal{L}} \; &(1,0)\, z(\tau)/J(\tau) - (1,1)\, z(\tau p)/J(\tau p) - (1,0)\, z(\nu)/J(\nu) + (1,1)\, z(\nu p)/J(\nu p)
\end{aligned}$$

for τ on $[0,1]$ and a given $\nu \in [0,1]$, which completes the proof of Lemma A.3. ■

D Additional supplementary contents

Here we provide additional supplementary contents, including an alternative identification result for $H^{-1}(\cdot)$ (briefly mentioned in Section 4.2), and the LCC test and identification results for the single-period setting.

D.1 An alternative identification result for $H^{-1}(\cdot)$

Assumption D.1 *For some known $\theta_c \in [\underline{\theta}, \min_{j=1,2} \theta_t^{j*}]$, the following holds: $\psi_1'(e_c) = \psi_2'(e_c)$, $\psi_1'(e) > \psi_2'(e)$ for all $e > e_c$, and $\psi_1'(e) < \psi_2'(e)$ for all $e < e_c$, with $e_c \equiv \max\{e(\theta_c;\varpi_1), e(\theta_c;\varpi_2)\}$.*

Corollary D.1 *Consider the single-period setting. Under Assumptions 2.1-2, 4.2(i) and D.1,* $H^{-1}(\cdot)$ *is identified on* $\left[\max_{j=1,2} c\left(\underline{\theta};\varpi_j\right), \min_{j=1,2} \bar{c}_t^j\right]$.

Proof of Corollary D.1. The proof of Corollary D.1 is similar to that of Proposition 4.2. The main difference of identification in Corollary D.1 is to first identify the intersection point e_c which corresponds to the intersection point of the two cost distributions under $W=\varpi_1$ and $W=\varpi_2$. Suppose without loss of generality that $\theta_t^{1*}\leq\theta_t^{2*}$. Consequently, $\min_{j=1,2}\theta_t^{j*}=\theta_t^{1*}$. Under Assumption D.1, it can be shown that there exists a $\theta_c\in[\underline{\theta},\theta^{1*}]$ such that $E\left(\tau_c/p_{t,1};\varpi_1\right) = E\left(\tau_c/p_{t,2};\varpi_2\right)=e_c$, where τ_c satisfies $\theta_c=\theta\left(\tau_c\right)$ due to the one-to-one mapping between cost and type. And, τ_c is identified by $C_t\left(\tau_c/p_{t,1};\varpi_1\right)=C_t\left(\tau_c/p_{t,2};\varpi_2\right)$ since the two distribution functions of $C_t\left(\cdot;\varpi_1\right)$ and $C_t\left(\cdot;\varpi_2\right)$ intersect only once at the cost quantile corresponding to $\theta\left(\tau_c\right)$. Consequently, e_c is identified. As a result, for $c\in\left[\max_{j=1,2} c\left(\underline{\theta};\varpi_j\right), \min_{j=1,2} \bar{c}_t^j\right]$, $H^{-1}(c)$ is identified up to $\theta\left(\tau_c\right)$ as

$$H^{-1}(c) \quad = \quad \theta(\tau_c)-e_c+\sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(c)\cdot p_{t,1}\right).$$

To identify $\theta\left(\tau_c\right)$, note that

$$H^{-1}\left(C_t(0;\varpi_1)\right) \quad = \quad \underline{\theta}-E_t\left(0,\varpi_1\right). \tag{D.1}$$

In addition, by following similar steps as in the proof of Proposition 4.2, it can be shown that

$$H^{-1}\left(C_t(0;\varpi_1)\right)=\theta(\tau_c)-e_c+\sum_{k=0}^{\infty}\Delta\tilde{E}\left(\tau_k(C_t(0;\varpi_1))\cdot p_{t,1}\right). \tag{D.2}$$

(D.1) and (D.2) together identify $\theta(\tau_c)$ as

$$\theta(\tau_c)=\underline{\theta}-E_t\left(0,\varpi_1\right)+e_c-\sum_{k=0}^{\infty}\Delta\tilde{E}_t\left(\tau_k(C_t(0;\varpi_1))\cdot p_{t,1}\right).$$

■

D.2 The LCC test and identification results for the single-period setting

Following essentially the same logic, the LCC test and identification strategies (without LCC) proposed in Sections 3 and 4 work for the single-period setting with only slight modifications needed, as we clarify in Appendix D.2.

Under the single-period setting, we observe the realized cost C, the payment Q, the exclusion variable W, the two effort proxies X and Y, and a dummy variable D^F that indicates the choice between FF and CR, with $D^F = 1$ indicating FF, and $D^F = 0$ indicating CR. Consequently, we have a random sample

$$\left\{c_i, q_i, w_i, x_i, y_i, d_i^F\right\}_{i=1}^{n}.$$

Further, define

$$\left\{c_i, q_i, w_i, x_i, y_i, d_i^F\right\}_{i\in\mathcal{FP}} \text{ with } \mathcal{FP} = \left\{i : d_i^F = 1\right\},$$

i.e., the subsample of FP contracts, which can be further divided into two groups by conditioning on W:

$$\mathcal{FP}^j \equiv \left\{i \in \mathcal{FP} : w_i = \varpi_j\right\}, \text{ for } j = 1, 2.$$

Econometric analyses of the single-period setting share the following notations with those of the two-period setting: $c(\theta; \varpi_j)$, $e(\theta; \varpi_j)$, $\psi_j(\cdot)$, for $j = 1, 2$, and $\theta(\tau)$ for $\tau \in [0, 1]$. We need to introduce a few new notations exclusive for the single-period settings, as follows: θ^{j*} denotes the cutoff value of θ between FP and CR conditional on $W = \varpi_j$. (For more details on θ^{j*}, please refer to Corollary 3.1 (i)); $E(\tau; \varpi_j)$ denotes the τ'th quantile of an effort distribution generated by the transformation $e(\cdot, \varpi_j)$ of the truncated distribution of θ on $\left[\underline{\theta}, \theta^{j*}\right]$. In other words, $E(\tau; \varpi_j)$ is the τ'th quantile of the effort conditional on FP being chosen and $W = \varpi_j$; Similarly, $C(\tau; \varpi_j)$ denotes the τ'th quantile of the cost conditional on FP being chosen and $W = \varpi_j$. Obviously, these new notations for the single-period setting parallel, and are simplified (by dropping subscript "t") from, those for the two-period setting.

For the single-period setting, Assumptions 3.1 and 4.1 are replaced by their single-period version, namely, Assumptions D.2 and D.3, respectively, below.

Assumption D.2 *The following conditions hold for the single-period setting (or for the two period-setting): (i) $\left\{c_i, q_i, w_i, d_i^F\right\}_{i=1}^{n}$ are independent and identically distributed across i for each n; (ii) The density $f(\cdot)$ of θ exists, and is bounded and continuous on a bounded support; (iii) The conditional density $f_{C|W}(c|\varpi_j)$ of cost c under FP for $j = 1, 2$ exist, and are bounded and uniformly continuous in c on bounded supports.*

Assumption D.3 *There exist two effort-related proxy variables X and Y such that*

$$X = E + V_1 \quad \text{and} \quad Y = m(E) + V_2,$$

for some unknown function $m(\cdot)$, where E, V_1, and V_2 are mutually independent with $\mathbb{E}(V_1) = \mathbb{E}(V_2) = 0$.

Condition T.1, Assumptions 3.2 and 4.2 remain the same for the single-period setting.

Under Assumption 3.2, the results in Propositions 2.1 hold conditioned on W. And it is helpful to (re)state these results as the following corollary:

Corollary D.2 *Let Assumption 3.2 hold. Conditioned on $W = \varpi_j$, the followings hold in equilibrium for the single-period setting: There is a cutoff value $\theta^{j*} \in [\underline{\theta}, \bar{\theta}]$ such that the agent chooses FP if $\theta \leq \theta^{j*}$, and chooses CR otherwise. And the optimal fixed price b^{j*} and θ^{j*} satisfy*

$$b^{j*} = H\left(\theta^{j*} - e\left(\theta^{j*}; \varpi_j\right)\right) + \psi_j\left(e\left(\theta^{j*}; \varpi_j\right)\right), \tag{D.3}$$

$$\left(1 - \frac{\alpha}{1+\lambda}\right) \frac{F\left(\theta^{j*}\right)}{f\left(\theta^{j*}\right)} = \frac{H\left(\theta^{j*}\right) - b^{j*}}{H'\left(\theta^{j*} - e\left(\theta^{j*}; \varpi_j\right)\right)}. \tag{D.4}$$

Clearly, Corollary D.2 parallels Propositions 2.1.

D.2.1 The LCC test

For the single-period setting, we test the null

$$\mathbb{H}_0: \ C\left(\tau; \varpi_1\right) - C\left(\tau p; \varpi_2\right) \equiv \text{a constant, for } \tau \in [0,1] \text{ almost sure },$$

which necessarily holds under the LCC. And we construct the test statistic as

$$T_n \equiv \int_0^1 n_f \left\{ \left[\widehat{C}\left(\tau; \varpi_1\right) - \widehat{C}\left(\tau \hat{p}; \varpi_2\right)\right] - \left[\widehat{C}\left(\nu; \varpi_1\right) - \widehat{C}\left(\nu \hat{p}; \varpi_2\right)\right]\right\}^2 d\tau \tag{D.5}$$

where: (i) $n_f \equiv \sum_{i=1}^n d_i^F$; (ii) $\hat{p} = \hat{p}_1/\hat{p}_2$, with $\hat{p}_j \equiv \sum_{i=1}^n d_i^F \mathbf{1}\left(w_i = \varpi_j\right) / \sum_{i=1}^n \mathbf{1}\left(w_i = \varpi_j\right)$. (iii) $\left\{\widehat{C}\left(\tau, \varpi_1\right), \widehat{C}\left(\tau \hat{p}, \varpi_2\right)\right\}$ are obtained from quantile regressions based on the subsample $\mathcal{FP}$. Specifically, $\widehat{C}\left(\tau, \varpi_1\right) = \widehat{q}_1\left(\tau\right)$ and $\widehat{C}\left(\tau \hat{p}; \varpi_2\right) = \widehat{q}_1\left(\tau \hat{p}\right) + \widehat{q}_2\left(\tau \hat{p}\right)$, where

$$\begin{aligned}
\left\{\widehat{q}_1\left(\tau\right), \widehat{q}_2\left(\tau\right)\right\} &= \underset{q_1,\, q_2}{\operatorname{argmin}} \sum_{i=1}^{n_f} \rho_\tau\left(c_{f_i} - q_1 - q_2 \cdot \mathbf{1}\left(w_{f_i} = \varpi_2\right)\right), \\
\left\{\widehat{q}_1\left(\tau \hat{p}\right), \widehat{q}_2\left(\tau \hat{p}\right)\right\} &= \underset{q_1,\, q_2}{\operatorname{argmin}} \sum_{i=1}^{n_f} \rho_{\tau \hat{p}}\left(c_{f_i} - q_1 - q_2 \cdot \mathbf{1}\left(w_{f_i} = \varpi_2\right)\right),
\end{aligned}$$

Under Condition T.1, Assumptions 3.2 and D.2, based on the test statistic T_n above and bootstrap critical values, $\mathbb{H}_0$ can be consistently tested.

D.2.2 Identification

Denote by $F_{E|W}\left(\cdot|\varpi_j\right)$ and $F_{C|W}\left(\cdot|\varpi_j\right)$ the conditional CDF of effort (under FP) and cost (under FP), respectively, on $W = \varpi_j$. Again, it follows from Theorem 1 of Schennach and Hu (2013) that $m(\cdot)$ and$F_{E|W}\left(\cdot|\varpi_j\right)$ are nonparametrically identifiable from the conditional distribution of (X, Y) on $W = \varpi_j$, for $j = 1, 2$. And $e = e\left(\theta; \varpi_j\right)$ can be recovered as

$$e = F_{E|W}^{-1}\left(F_{C|W}(c|\varpi_j)|\varpi_j\right), \ \forall \, c \in [\underline{c}^j, \bar{c}^j] \equiv \left[c\left(\underline{\theta}; \varpi_j\right), c\left(\theta^{j*}; \varpi_j\right)\right], \tag{D.6}$$

The cost function and type distribution

Following essentially the same steps as in Sections 4.2 and 4.3, under Assumptions 3.2, 4.2, D.2 and D.3, $H^{-1}(\cdot)$ is nonparametrically identified on $[\underline{c}, \bar{c}^1]$ as

$$H^{-1}(c) = \begin{cases} \underline{\theta} - \underline{e}, & \text{for } c = \underline{c}, \\ \underline{\theta} - \underline{e} + \sum_{k=0}^{\infty} \Delta\tilde{E}(\tau_k(c) \cdot p_1), & \text{for } c \in \left(\underline{c}, \bar{c}^1\right], \end{cases} \tag{D.7}$$

where $\Delta\tilde{E}(\tau) \equiv E(\tau/p_2; \varpi_2) - E(\tau/p_1; \varpi_1)$ for any $\tau \in [0, p_1]$, and $p_j = F\left(\theta^{j*}\right) = \mathbb{E}\left(D^F | W = \varpi_j\right)$.

Based on identification of $H^{-1}(\cdot)$ on $\left[\underline{c}, \bar{c}^1\right]$, we can recover the corresponding θ for any $c \in \left[\underline{c}, \bar{c}^1\right]$ conditioned on $W = \varpi_1$ as

$$\theta = H^{-1}(c) + F_{E|W}^{-1}\left(F_{C|W}(c|\varpi_1)|\varpi_1\right), \forall c \in \left[\underline{c}, \bar{c}_t^1\right]. \tag{D.8}$$

Consequently, we can identify $G(\cdot)$ and $g(\cdot)$, the truncated CDF and pdf of θ on $\left[\underline{\theta}, \theta^{1*}\right]$. We can then identify the (unconditional) CDF and pdf of θ on $\left[\underline{\theta}, \theta^{1*}\right]$ as $F(\cdot) = G(\cdot)F\left(\theta^{1*}\right)$ and $f(\cdot) = g(\cdot)F\left(\theta^{1*}\right)$, respectively, noting that $F\left(\theta^{1*}\right) = \mathbb{E}\left(D^F | W = \varpi_1\right)$ is readily identifiable.

The disutility functions

It follows from (D.6) that, conditioned on $W = \varpi_j$, the cost c corresponding to any effort $e \in \left[\underline{e}, \bar{e}^j\right]$ is given by $c = F_{C|W}^{-1}\left(F_{E|W}(e|\varpi_j)|\varpi_j\right) \in \left[\underline{c}, \bar{c}^j\right]$, which, together with the f.o.c. (3.3), imply that, for any $e \in \left[\underline{e}, \bar{e}^j\right]$,

$$\psi_j'(e) = H'\left(H^{-1}\left(F_{C|W}^{-1}\left(F_{E|W}(e|\varpi_j)|\varpi_j\right)\right)\right) = \frac{1}{H^{-1\prime}\left(F_{C|W}^{-1}\left(F_{E|W}(e|\varpi_j)|\varpi_j\right)\right)} \tag{D.9}$$

Given identification results already established, $\psi_j(\cdot)$ is identified by its differential equation (D.9), together with the location normalization condition $\psi_j(0) = 0$ by Assumption 3.2 (ii).

The ratio $\alpha/(1+\lambda)$

Rearranging terms for Eq (D.4) yields

$$\begin{aligned} \frac{\alpha}{1+\lambda} &= 1 - \frac{H\left(\theta^{1*}\right) - b^{1*}}{H'\left(\theta^{1*} - e\left(\theta^{1*}; \varpi_1\right)\right)} \frac{f\left(\theta^{1*}\right)}{F\left(\theta^{1*}\right)} \\ &= 1 - \frac{\underline{c}_{CR}^1 - b^{1*}}{H'\left(H^{-1}\left(\bar{c}^1\right)\right)} \frac{f\left(\theta^{1*}\right)}{\mathbb{E}\left(D^F | W = \varpi_1\right)}, \end{aligned} \tag{D.10}$$

with $\underline{c}_{CR}^1 \equiv H\left(\theta^{1*}\right) \left(> H\left(\theta^{1*} - \bar{e}^1\right) = \bar{c}^1\right)$ being the lower bound for cost under CR, conditioned on $W = \varpi_1$. Since all terms of (D.10) are either directly identifiable from the observed variables or already shown identifiable, (D.10) identifies $\alpha/(1+\lambda)$.

To summarize, under Assumptions 3.2, 4.2, D.2 and D.3, $H^{-1}(\cdot)$ is nonparametrically identified on $\left[\underline{c}, \bar{c}^1\right]$. Consequently, $H(\cdot)$, $F(\cdot)$ and $(\psi_1(\cdot), \psi_2(\cdot))'$ are nonparametrically identified on $\left[\underline{\theta} - \underline{e}, \theta^{1*} - \bar{e}^1\right]$, $\left[\underline{\theta}, \theta^{1*}\right]$ and $\left[\underline{e}, \bar{e}^1\right]$, respectively. In addition, $\alpha/(1+\lambda)$ is identified. Note that γ (indicating intertemporal preference) is not involved in the single-period setting.