Distinguishing Incentive from Selection Effects in Auction-Determined Contracts *

Laurent Lamy†  Manasa Patnam ‡  Michael Visser§

January 18, 2019

Abstract

This paper develops a novel approach to estimate the incentive effects of contracts when the matching between agents and principals, and (part of) the contract features, derive from an auction process. The post-auction outcome of each agent depends on the principal-agent observable characteristics, the contract features and also on unobserved signals which drive how principals bid for the agents. We propose a control-function approach to account for the endogeneity of contracts and matching. This consists of, first, estimating the primitives of an interdependent values auction model- which is shown to be non-parametrically identified from the bidding data - second, constructing two control function terms that capture the expected signals of the auction winner and of the incumbent (i.e., the principal previously matched with the agent) and, third, plugging the estimated control terms into the post-auction outcome equation. A Monte Carlo study shows that the extent of the biases can be large when the endogeneity problem is ignored, but that estimation of our augmented outcome equation leads to satisfactory results, even in small samples. We apply our methodology to a labor market application: we estimate the effect of sports players’ auction-determined wages on their individual performance.

Keywords: Sample Selection; Econometrics of Contracts; Econometrics of Auctions; Endogenous Matching; Wage-Performance Elasticity.

JEL classification: C01 ; C29 ; C57 ; D44; M52; Z22.

*We thank James Albrecht, Shekar Shankar Aiyar, Mitali Das, Kartikeya Date, Xavier d’Haultfoeuille, Aureo de Paula, Guido Kuersteiner, Ingmar Prucha, Paul Scott, Ali Skalli, Andrew Sweeting, Frank Vella and seminar participants at CREST, Georgetown University, the University of Besançon, the University of Luxembourg, the University of Maryland, the University of Paris II, the 2nd ECODEC Conference, and the 7th French Econometrics Conference for helpful comments and suggestions. We thank Arun Kumar for excellent research assistance. This paper is a major revision of a draft which circulated under the title “Correcting for Sample Selection From Competitive Bidding, with an Application to Estimating the Effect of Wages on Performance.” Financial support from the LABEX ECODEC (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) is gratefully acknowledged.

†CIRED (Ecole des Ponts, ParisTech). Email: laurent.lamy78@gmail.com.
‡IMF. Email: mpatnam1@gmail.com.
§CREST-LEI and CRED, University of Paris 2. Email: michael.visser@ensae.fr.
1 Introduction

A central issue in the empirical literature on contracts and contracting behavior concerns estimating the effects of moral hazard (contracts offer a variety of incentives to elicit ex-post efforts from agents) and of endogenous selection (agents’ characteristics drive how agents are matched to contracts). The assessment of the relative magnitude between these effects is crucial when deriving either optimal contracting choices or optimal market regulation (Laffont and Tirole (1993)). Yet, a pervasive problem is that the empirical correlations between contracts features and ex post outcomes are not insightful because the endogenous matching between agents and contracts may be related to unobserved heterogeneity which drive also their ex-post outcomes (Chiappori and Salanié (2003)). The measurement of incentives effects is thus possibly confounded by the presence of selection effects and omitted variables. A large part of the empirical literature tackles these issues by exploiting exogenous variations in contract choices, by way of field or natural experiments (see e.g. Lazear (2000) and Shearer (2004)). Other recent approaches to model the match between agents and contracts consist in assuming that the menu of contracts is optimally set by a monopolistic principal and that agents pick the contract they prefer.\(^1\)\(^2\)

These approaches, however, abstract from the issue of endogenous principal-agent matching and cannot, as a result, identify some selection effects separately, limiting the assessment of optimal contract choice in environments with many principals. These approaches, however, rely crucially on the assumption that the matching between the principal and the agents is exogenous, e.g. because the principal has a monopoly position (which could be relevant on a short-term basis). They abstract thus from the issue that part of the selection effects comes from endogenous principal-agent matching: if the principal changes the contractual incentives, then it would change the set of agents he is matched with (especially on a long-term basis). A separate strand of the literature that relies exclusively on matching data explicitly analyzes the matching determinants, and by doing so, is able to offer some insights about incentive and selection effects, but only indirectly as ex-post outcomes are not observed and modeled (Ackerberg and Botticini (2002)).\(^3\) To the best of our knowledge, there is no empirical work

\(^1\)Relying heavily on contract theory, Perrigne and Vuong (2011) show how the first-order conditions of both the principal and the agents allow to identify non-parametrically the distribution of the agents’ unobserved characteristics and their demand function. Vuong, Luo, and Perrigne (in press, forthcoming) develop such a strategy to build a non-parametric estimator that is applied to cellular phone consumption data. See also Dubois and Vukina (2009) for an analysis of grower contracts with a full parametric perspective.

\(^2\)d’Haultfoeuille and Février (2011) is related to both approaches: benefiting from a natural experiment in the remuneration rule of agents making household surveys, they identify non-parametrically the distribution of the interviewers’ cost of effort, which allows then to develop various counterfactual, including the gains from switching to the optimal menu of contracts.

\(^3\)In contrast with previous works on agricultural contracts, Ackerberg and Botticini (2002) found evidence of risk sharing once endogenous matching is accounted for (see also Kandilov and Vukina (2016)). See Chiappori and Salanié (2016) for a survey on the structural econometrics of frictionless matching models that rely on some stability conditions, an approach that is popular from Choo and Siow (2006) to model marriage markets with unobserved
that estimates the incentive effects of contracts from data where principals and agents have unobserved characteristics that drive how they are matched, i.e. when both matching and contract choices are endogenous and a post-contract outcome is observed.

At the core of the matching decisions and the source of possible endogeneity they cause, various forces play a role among which 1) principal-agent idiosyncratic synergies and 2) asymmetric information about common value elements. For instance in labor market settings, the context of our application, the wage offered by a firm to a worker depends not only on the level of synergy between them (such as the match between the worker's skill and the job) but also on certain worker attributes that are commonly valued across all firms (for example, the worker's ability). Firms that were previously matched with a given worker benefit from superior information on such common value attributes which can significantly affect worker mobility and the distribution of equilibrium wages (Greenwald (1986), Pinkston (2009)).

This paper develops a methodology to consistently estimate the effects of contract features and principal-agent characteristics in the presence of unobserved heterogeneity that drives the matching of principals to agents and determines the contractual terms. Our methodology consists in modeling this endogeneity as the outcome of an auction model, where principals value both the expected ex-post outcome related to the agent as well as the agent's observable characteristics. The principals receive private signals that are informative about idiosyncratic synergies with the various agents. In addition, principals who were previously matched to a given agent (referred to throughout as the ‘incumbent’) receive a private signal about attributes which are commonly valued among the various principals. Our auction model departs thus from pure private values models and belongs rather to the larger class of interdependent values auction models with asymmetric bidders (Krishna (2002)). Contract features are determined by the auction and may have a causal ‘incentive’ effect on the agent’s post-contract outcome which we assume principals to perfectly anticipate when submitting their bids and which thus induces a departure from quasi-linearity.

Our first step, which relies on bidding data alone, consists in non-parametrically identifying a bidding model and in particular the distribution of some underlying bid-signals that drive principals' bidding behavior. This step is anchored both in auction theory and in the literature on the structural econometrics of auction data (see Athey and Haile (2008) for a survey on identification), the most novel aspect of our analysis being on the econometric side given that the literature is typically limited to pure private values since interdependent values models are known to cause important identification problems (Laffont and Vuong (1996)). Equipped with an estimate of the principals’ bid-signals distribution conditional on the outcome of the auction market, we then identify and build two control terms that allow us to tackle the various sources of endogeneity. The first control term is associated with the idiosyncratic synergies between

---

4See Oyer and Schaefer (2010) for a comprehensive survey on hiring and incentives in the labor market.
the agent and the principal he is matched with, while the second control term captures agent-specific heterogeneity that is observed solely by the incumbent. Our control function approach is related to the two-step estimators that have been developed in the sample selection literature pioneered by Heckman (1979) for binary outcomes and later extended to polychotomous outcomes by Lee (1983) or Dubin and McFadden (1984).\footnote{See Vella (1998) and Wooldridge (2015) for surveys. Wooldridge (2002) and Das et al. (2003) also propose a control function method to address these two issues simultaneously with a parametric and non-parametric approach, respectively. Unfortunately the setup considered by those authors is not easily generalized to allow for selection on multiple outcomes, and to take into account the particular way in which our sample selection and endogeneity arises. More generally, this literature is no longer rooted into structural matching/selection models unlike Dubin and McFadden (1984) which was rooted into discrete choice theory.}

In a similar vein, Cassola, Hortaçsu, and Kastl (2013) develop a two-stage approach where the first stage consists of recovering bank’ private values from their bid in a multi-unit discriminatory auction\footnote{See Hortaçsu and McAdams (2018) for a survey on the econometrics of multi-object auctions under the private value paradigm.} and the second stage establishes correlations between banks’ auction valuations (their willingness-to-pay for liquidity) and some ex-post measures (e.g. banks balance sheet). The main novelty of our work is that we develop a model where the auction stage and the ex-post outcome equation are structurally linked, through some signals observed privately by the bidders, and where the auction outcome may have a causal impact on the ex-post outcome (what we refer to as the incentive or moral hazard effect). Nevertheless, we also stress that, to the best of our knowledge, our structural approach is novel even in the simple case where the moral hazard effect is absent or if the auction model is a pure private values model. Even under such simple environments, the measurement of the effects of observable covariates is confounded by the presence of unobserved heterogeneity. Indeed, results from our Monte Carlo simulations reveal that there is a large bias in the effects not only of incentives features but also of observable covariates from ignoring contract endogeneity; furthermore the extent of those biases grows with the strength of common values present in the auction.

Overall, our approach integrates different strands of literature. It not only builds on the existing literature on sample selection and the structural econometrics of auctions, but also offers novel insights into the treatment of endogenous matching with multiple outcomes and the non-parametric identification of an auction model involving interdependent values, asymmetries between bidders and where bidders’ payoff functions are no longer quasi-linear in the auction price. In order to disentangle idiosyncratic synergies from common value components in the bid of an incumbent, we assume that the associated signals are drawn independently and we also make an exclusion restriction by assuming that the signal distributions are the same in the samples with and without an incumbent.\footnote{Relatedly, the independence restriction is commonly used to deal with unobserved heterogeneity (Li, Perrigne, and Vuong (2000) and Krasnokutskaya (2011)) while similar exclusion restrictions -relying on exogenous variations on the intensity of competition- are used to test for common values (Haile, Hong, and Shum (2003)), to identify risk-aversion (Lu and Perrigne (2008)) or to allow for correlated private values (Aradillas-Lopez, Gandhi, and Quint (2013)).} In doing so, from a more applied perspective,
we contribute to the literature on empirical contracts by offering a consistent estimator that allows to test key preoccupations in contract theory.

Our framework can be applied very broadly, especially in environments where an analyst is interested in the relationship between a post-auction outcome and the auction price.\(^8\) One of the most relevant applications we have in mind are procurement contracts. In particular, most papers on procurement auctions\(^9\) abstract from the crucial aspect that the monetary transfers between the contractor and the contractee depend on the performance of the realization. Another promising application is in the area of auctions for renewable energies where firms' bids correspond typically to a subsidy per quantity produced. Given the huge discrepancies in terms of production performance across countries (Huenteler et al. (2018)), this raises economically important issues for the issue of efficient deployment of renewable energies.

In our application we focus on labor market auctions. This is an interesting application for our methodology, first, because they are gaining empirical relevance thanks to increased digitization\(^10\) and, second, because auctions serve as a popular modeling tool in labor economics.\(^11\) We estimate the relationship between wages and performance of sports players. The wages in our data set are determined through a sequence of English auctions with some reserve price. In these auctions the bidders are the teams participating in a cricket tournament and the ‘objects’ for sale are the cricket players with the winning price represents the cricketer’s wage. Our results indicate a downward bias in the positive and statistically significant wage effect, after accounting for selection and endogeneity using our control function terms derived from our first stage auction model. The coefficients on the incumbent’s control function is positive and significant, suggesting that selection effects play a key role. Our findings from this application contributes to the strand of literature which tests for the presence of reciprocity (Akerlof 1982) or fairness (Akerlof and Yellen 1990) effects.\(^12\)

\(^8\)Very few papers consider the joint modeling of auction and post-auction data. Notable exceptions are Athey and Levin (2001) and Hendricks, Pinkse, and Porter (2003). However, these papers have a different objective than ours and aim more to leverage post-auction data to deal with more general information structures at the auction stage. Even closer to our perspective, Bhattacharya, Ordin, and Roberts (2018) estimate a model for oil and gas auctions where the timing of an ex-post investment (drilling) depends on the realization of partially observed variables that are uncertain at the bidding stage. The model is estimated from an auction setup where moral hazard is not explicitly modeled. However, moral hazard is indirectly present in some auction designs that are evaluated via counterfactuals. Another difference is that Bhattacharya, Ordin, and Roberts (2018) consider only a pure common value symmetric model.

\(^9\)Two notable empirical exceptions are Bajari et al. (2014) and Lewis and Bajari (2014) who highlight the first order importance of incentives in procurements for highway maintenance (regarding renegotiation and completion time respectively).

\(^10\)For example, labor market auctions proliferate on the Internet, where employers can post job ads and solicit bids from interested workers. See bit.ly/OnlineLaborAuction for examples of online labor markets where employers and job-seekers are matched through auctions.

\(^11\)Interdependent values model can serve as a key ingredient to explain such environments. See for example, Pinkston (2009) for a pure common-value auction model which presents similarities with our model but with a completely different empirical perspective.

\(^12\)So far the studies in this branch of the literature have been based primarily on lab or field experiments. The seminal contribution is Fehr, Kirchsteiger, and Riedl (1993) whose results provide support for the reciprocity hypothesis: offering wages above the market-clearing wage elicit higher efforts. Lee and Rupp (2007) exploit the fact
The remainder of the paper is organized as follows. Section 2 presents the environment and how the omission of match-relevant private signals can bias OLS estimates. Section 3 characterizes how firms and workers are matched as the equilibrium of an interdependent auction model. Section 4 shows that the bidding model is non-parametrically identified from the bidding data alone, which enables us to identify the control functions. Section 5 relates our contribution to the literature. Section 6 presents the results of some Monte Carlo simulations. Section 7 describes the labor market environment used for our application and resulting estimates. Section 8 concludes.

2 The environment and the endogeneity problem

We consider a (small) collection of firms, indexed by \( f = 1, \ldots, F \), bidding for a (large) collection of workers, indexed by \( i = 1, \ldots, N \). The workers are auctioned sequentially, one after another. An observable reserve price, denoted by \( W_r^i > 0 \), is attached to each worker \( i \) before the auction sequence starts: it corresponds to the starting price in the auction for \( i \). The winner of this auction (if there is one) is denoted \( f_w^i \), and the wage at which \( i \) is employed corresponds to the (observable) auction price, and is denoted \( w_i \geq W_r^i \). The bidding rules are detailed in the next section and the associated equilibrium will then determine both the auction price and the auction winner as a function of bidders’ private information. We take into account the possibility that \( i \) worked in one of the firms prior to the auctions. In such a case, the corresponding firm is referred to as the incumbent and denoted by \( f_{inc}^i \in \{1, \ldots, F\} \). Otherwise (i.e., if worker \( i \) was unemployed before or employed by a firm not belonging to the set of potential participants for auction \( i \)), we let \( f_{inc}^i := 0 \).

Before the auction for \( i \) starts, each potential participant \( f \) receives a single-dimensional private signal, denoted by \( s_{PV}^{i, f} \in \mathbb{R} \). This signal is only observed by \( f \) and summarizes the synergies between the firm and \( i \) regarding the worker’s future performance. The incumbent (if any) receives also an additional single-dimensional private signal, denoted by \( s_{CV}^{i} \in \mathbb{R} \). This signal is only observed by the incumbent and summarizes \( i \)'s performance capabilities that are relevant not only if \( i \) works for the incumbent, but also if this worker is employed by another firm. It captures determinants which were revealed to the incumbent through interaction with the worker on the workplace (learning by doing). The incumbent thus has superior information as in Engelbrecht-Wiggans et al.’s (1983) pure common value auction model with an informed bidder. Additionally, we consider a framework where the wage itself may have a causal effect on the worker’s future performance. Formally, the performance of worker \( i \) if employed by firm

---

13 Beyond the auction literature, the distortion resulting from an incumbency status has received some attention in the labor literature, and in particular whether informational problem can impede employer-to-employer worker flows. See Pinkston (2009) for a bidding model under asymmetric employer learning.
\(y_{i,f} = \beta_f + \beta_x \cdot x_{i,f} + \tau \cdot \log(w_i) + s_i^{PV} + s_i^{CV} + \epsilon_{i,f}\)

(1)

Here \(\beta_f\) represents a firm-specific fixed effect (skills of the managerial staff; the firm’s equipment and infrastructure), \(x_{i,f}\) a vector of observable characteristics of both \(i\) and \(f\) that are publicly observable at the auction stage, and \(\beta_x\) a vector of parameters measuring the effects of the corresponding elements in \(x_{i,f}\). We assume that the vector \(x_{i,f}\) includes whether firm \(f\) is the incumbent with respect to worker \(i\) through the dummy variable \(x_{i,f}^{dum}\), which is equal to 1 if \(f = f_{i,inc}\) and 0 otherwise, and we let \(\beta_x^{dum}\) denote the scalar in \(\beta_x\) that measures the corresponding effect. The variable \(x_{i,f}\) may include features of the contracts that are determined prior to the auction so that the corresponding element in \(\beta_x\) represents an incentive effect. The term \(\tau \cdot \log(w_i)\) captures the direct effect of wage on performance. It corresponds to one of the incentive effects of interest referred to in the introduction, i.e. the effect of the endogenous contractual feature (here the wage) on performance. However, all our analysis is also relevant when \(\tau = 0\), i.e. when the incentive effects are captured only by the exogenous variables \(x_{i,f}\). Finally, \(\epsilon_{i,f}\) captures other performance determinants that are unobserved by the firms (until worker \(i\) actually starts working). We assume that the vectors of performance shocks \(\epsilon_{i,f}, f = 1, \cdots, F\), are i.i.d. across all \(i\) and that their means are equal to zero. The signals \(s_i^{CV}\) and \(s_i^{PV}\) are also centered around zero (additional assumptions on these signals are introduced in Section 3). The econometrician neither observes these two signals nor the performance shock. From his perspective the error term in the performance equation is the sum of these three components, denoted \(u_{i,f}\).

To estimate \(\beta_x\) and \(\tau\), we face two kinds of problems. The first is that estimation is based on a selected sample. The selection arises because the wage and performance of a given worker \(i\) are only observed if some firms decide to participate, i.e. to enter a bid at or above the reserve price \(W_i^r\). In addition, if there is bidding above the reserve price, only recorded is the productivity \(y_{i,f}^{w}\), while \(y_{i,f}\) is naturally unknown and counterfactual for all \(f \neq f_i^{w}\). Intuitively, worker \(i\) being matched with firm \(f\) reveals that this firm is willing to pay more than the reserve price, and also values employing \(i\) more than his competitors. Given that we expect the incumbent firm \(f_{i,inc}\) (resp. a non-incumbent firm \(f \neq f_{i,inc}\)) to bid more aggressively the higher the term \(\beta_x \cdot x_{i,f}^{inc} + s_{i,f}^{PV} + s_i^{CV}\) (resp. \(\beta_x \cdot x_{i,f} + s_{i,f}^{PV}\)) – an intuition confirmed and formally shown in the next section – the sample selection is not random but related to \(x_{i,f}, s_{i,f}^{PV}\) for all \(f\), and to \(s_i^{CV}\). Belonging to the selected sample or not thus depends on the observed worker/firm characteristics, but also on the unobserved signals received by the firms. Both these types of

---

\(^{14}\)Our analysis extends straightforwardly to environments where the distribution of the residuals \(\epsilon\) depend on the observable covariates. In other words, we could allow the shocks \(\epsilon_{i,f}^{w}\) to be heteroskedastic w.r.t. the vector \(x_{i,d}^{w}\).
variables are at the same time determinants of performance. Estimation of the performance equation (1) by OLS leads to biased estimates because of the link between the selection rule and the composed error term $u_{i,f}$. More precisely, there is a bias in the OLS estimates because the mean of this error term conditional on being in the sample (and given the wage and the other observable explanatory variables), which is denoted by $E[u_{i,f} | w_i, x_{i,f}]$, is non-zero and varies across observations.\footnote{For notational simplicity we omit that the expectation of $u_{i,f}$ is also conditional on $f^*_{i}$ being the auction winner and conditional on $w_i \geq W^r_i$.} For example, if the incumbent wins the auction (i.e., $f^*_{i} = f^{inc}_{i}$), we have $E[u_{i,f_{i}^{w}} | w_i, x_{i,f_{i}^{w}}] = E[s^{PV}_{i,f_{i}^{w}} + s^{CV}_{i} | w_i, x_{i,f_{i}^{w}}]$, which is expected to decrease with $\beta_x \cdot x_{i,f_{i}^{w}}$ (if this term is small the winner only bids aggressively if the sum of the signals $s^{PV}_{i,f_{i}^{w}} + s^{CV}_{i}$ is large). If positive synergies with the incumbent get larger (formally, if $\beta^{dum}_{s} > 0$ increases), then it means that non-incumbents will need to have on average a larger private signal $s^{PV}_{i,f_{i}^{w}}$ in order to beat the incumbent. Due to sample selection and given that signals are not observed, the incumbency status will seem less attractive for performance that it really does. Finding no synergies between a worker $i$ and its incumbent $f^{inc}_{i}$ (or equivalently no switching cost from moving from one employer to another) with the OLS estimate of the incumbent dummy could thus merely be an artefact of this sample selection bias.\footnote{Under $\tau = 0$ and $\beta_s = (0, ..., \beta^{dum}_{s}, ..., 0)$, the OLS estimator of $\beta^{dum}_{s} > 0$ (resp. $\beta^{dum}_{s} < 0$) is downward (upward) biased if $E[u_{i,f_{i}^{w}} | w_i, x_{i,f_{i}^{w}}]$ decreases with $\beta^{dum}_{s} \cdot x_{i,f_{i}^{w}}$.}

The second problem is related to the first one, but is nonetheless distinct. It concerns the fact that $w_i$ is potentially endogenous in (1). Large values of $u_{i,f}$, for $f = 1, ..., F$, most likely indicate large values of $s^{CV}_{i}$ and $s^{PV}_{i,f}$, for $f = 1, ..., F$, which in turn should lead to a higher final price $w_i$. There is therefore potentially a positive relationship between the error term and the wage in the selected sample, so we suspect that $E[u_{i,f_{i}^{w}} | w_i, x_{i,f_{i}^{w}}]$ increases with $w_i$. A strong positive OLS estimate of the wage effect could thus merely be an artefact of this endogeneity bias.\footnote{Under $\beta_s = (0, ..., 0)$, the OLS estimator of $\tau$ is upward biased if $E[u_{i,f_{i}^{w}} | w_i, x_{i,f_{i}^{w}}]$ increases with $w_i$.}

The Monte Carlo simulations reported in Section 6 not only confirm our conjectures regarding the direction of the different biases, but also show that the naive OLS estimates tend to be far away from the true parameter values.

If the endogeneity of wages were the only source of bias (i.e., in the absence of sample selection),\footnote{Sample selection would not play a role in our model if reserve prices are either nonexistent or not binding, and if all workers are somehow randomly assigned to firms.} a standard instrumental variable method could be used to solve the problem provided a suitable instrument is available. Contrastingly, if selectivity were the only source of bias, i.e., when $\tau$ is known to be 0 and log($w_i$) drops out from the performance equation, then there
is still a sample selection problem with polychotomous outcomes, as in Dubin and McFadden (1984). Given that there are \( F + 1 \) possible outcomes in our setting (\( F \) possible winners plus the outcome with no sale), a pure IV approach seems unfeasible as it would require in our context \( F \) instrumental variables.

Our econometric strategy to control for the two sources of bias is therefore instead based on a control function approach as in Dubin and McFadden’s (1984) seminal contribution, which is particularly relevant in our case given that the auction literature is a mature field to rely on to build relevant control functions. Let \( \mathcal{S} \) denote the set of variables observed by the econometrician right at the end of the auction sequence. This set includes in particular the variables \( x_{i,f}, W_i \), and \( z_{i,f} \), for all \( i \) and \( f \), and \( w_i \) for workers \( i \) actually sold, but excludes the performance measures. Here \( z_{i,f} \) denotes a vector of variables observed by all bidders right before the auction for worker \( i \) starts. It includes covariates such as the order of sale of \( i \) in the auction sequence (whether this worker comes up for sale first, or second, etc...), the amount of money spent by \( f \) in the preceding auctions, the characteristics of the previous workers purchased by this firm, and possibly information on the purchases of its competitors. These variables are referred to as “the auction variables”. Relying on techniques from the structural econometric auction literature, we model and estimate the expectations \( E[s^{PV}_{i,f,w} | \mathcal{S}] \) and \( E[s^{CV}_{i} | \mathcal{S}] \) up to some multiplicative constant, and then add the estimates as control functions in the performance equations. While the auction variables \( z \) are not mandatory from a theoretical point of view, they will be helpful from an empirical perspective by playing the role of exogenous shifters in the control function terms, thereby reducing the collinearity between these control terms and the other variables appearing in the performance equations. As a prerequisite to deriving the precise form of the control functions, we now turn to the description of the auction model and the corresponding equilibrium.

3 The auction model

Here we develop an auction model in which firms compete with each other to buy the services of workers. Section 3.1 defines the auction rules. Section 3.2 defines the payoff derived by each firm from hiring a given worker as a function of the observable covariates, the private signals and the wage. Section 3.3 develops the equilibrium analysis, and Section 5 relates our model to the literature.

3.1 Auction rules

Workers are sold sequentially through English auctions with publicly known reserve prices. We formalize the game with the English button auction model (Milgrom and Weber (1982)). For worker \( i \), the incumbent (if any) first decides whether to enter the auction at the reserve
price $W^R_i$. Given this observable entry decision, non-incumbents then choose simultaneously whether to enter the auction themselves.\footnote{If we consider alternatively that the incumbent and the non-incumbents move simultaneously at the entry stage, then the equilibrium strategy of non-incumbents would be a bit different: for each non-incumbent, there would be a mass of signals such that he enters the auction but would then drop-out immediately if the incumbent has not entered. There would thus be ties with positive probability at the reserve price.} There are now three possibilities. First, if there are no entrants at all, the auction stops and the worker is not employed. Second, if there is a single entrant, the auction also stops but the worker is employed by this entrant and at the reserve price. Finally, if there are multiple entrants, the auction clock starts ticking at $W^R$, and moves up the price continuously. As the price $p$ goes up, entrants have to decide constantly whether to remain active in the auction, or to exit irrevocably. The clock stops when there is only one remaining active bidder. This bidder becomes the winner, and the wage $w$ paid to the worker corresponds to the auction termination/final price.\footnote{The tie-breaking rule (stipulating what happens when several bidders exit simultaneously at the termination price) is left unspecified since it does not play any role in our equilibrium analysis.} Furthermore, at any price reached by the clock, we assume that non-incumbents observe perfectly whether the incumbent is still active or not.

### 3.2 Bidders’ preferences and cutoff bid-signals

We assume that firms’ preferences over workers do not depend solely on the performance signals introduced in the previous section, but also possibly on additional signals. This reflects that workers are valued by firms for possibly many other reasons than the performance measure we use.

The information-structure is analogous to the one adopted in the previous section: for any given worker $i$, each firm $f$ is assumed to receive a private signal, denoted $s_{id}^{i,f} \in \mathbb{R}$, and that enters only the payoff function of firm $f$ when winning the auction; the incumbent (if any) receives an additional private signal, denoted $s_{i}^{co} \in \mathbb{R}$, and that enters the payoff function of the winning bidder independently of its identity. As for the signal $s_{i}^{CV}$, the signal $s_{i}^{co}$ reflects insider information held by the incumbent thanks to the previous relationship with the worker. The signals $s_{i,f}^{id}$, $s_{i,f}^{PV}$ are normalized such that their mean is zero. We let $s_{i,f}$ denote the vector of firm $f$’s private signals, i.e., $s_{i,f}^{inc} = (s_{id}^{i,f}, \ldots, s_{i,f}^{PV})$ for the incumbent and $s_{i,f} = (s_{id}^{i,f}, s_{i,f}^{PV})$ if $f \neq f_{inc}$. Throughout our analysis we make the following two independence assumptions.

**A1:** i) The vector $s_i := ((s_{i,f}^{id}, s_{i,f}^{PV})_{f=1,\ldots,F}, s_{i}^{co}, s_{i}^{CV})$ is i.i.d. across $i$. ii) For each $i$, the signals $s_{i,f}$ are distributed independently across $f$.

The first assumption stated in A1 is standard in the empirical auction literature, and ensures that our estimators have the usual asymptotic properties. For simplicity, it also leads us to drop the worker index $i$ from our notation when it is superfluous. The second assumption is required
for the auction model to be non-parametrically identified. It also matters for our equilibrium characterization. Note that ii) does not impose $s_{i,f}$ to be identically distributed across $f$, nor does it restrict the various signals belonging to $s_{i,f}$ to be independent.  

Our analysis captures the sequential/dynamic aspect of the game by including some auction covariates $z_{i,f}$ when defining below bidder $f$’s valuation $V_{i,f}$: the vector $z_{i,f}$ can include e.g. the amount of money spent by $f$ in the auctions preceding the auction for worker $i$, the number and type of workers bought prior to $i$ (capturing possible substitutabilities or complementarities between employees), and so forth. Instead of defining firms’ preferences for any bundle of workers, we define firms’ preferences for any given worker and any given auction covariates. This is a reduced-form approach to capture continuation values as developed by Jofre-Bonet and Pesendorfer (2003) in a dynamic game perspective. In empirical works on procurements data, it is now standard to use backlog variables as covariates. From a strategic perspective, it allows us to work as if each worker is sold in isolation. More generally, the auction covariates capture variables that enter the valuation equation but not the performance equation and that will play thus the role of an instrument.

We assume that firms are risk-neutral and hence maximize their expected payoff. We furthermore assume that the payoff derived by firm $f$ from winning the auction for worker $i$ at wage $w$ is given by

$$V_{i,f} - w \equiv \bar{V}_f(x_{i,f}, z_{i,f}) \cdot e^{\lambda [s_{i,f}^{id} + s_{i,f}^{co}]} \cdot E_{\epsilon_{i,f}}[e^{\lambda \cdot y_{i,f}}] - w,$$

while the payoff of each losing bidder is normalized to zero.  

The payoff of firm $f$ is the difference between the valuation of employing worker $i$, denoted $V_{i,f}$, and the wage $w$ that the worker is going to receive from this firm. The valuation is the product of three terms: \(\bar{V}_f(x_{i,f}, z_{i,f})\), which depends only on observable covariates; \(e^{\lambda [s_{i,f}^{id} + s_{i,f}^{co}]}\), which depends on the signals $s_{i,f}^{id}$ and $s_{i,f}^{co}$; and the expectation (with respect to the performance shock $\epsilon_{i,f}$) of $e^{\lambda \cdot y_{i,f}}$. \(\bar{V}_f(., .)\) is a strictly positive function of the firm/worker characteristics $x_{i,f}$ and the auction variables $z_{i,f}$. The firm/worker characteristics can thus have an indirect effect on firm $f$’s payoff through the performance $y_{i,f}$, and a direct effect through the function $\bar{V}_f$. The direct effect captures the possibility that $f$ values firm/worker characteristics intrinsically. The firm may, for instance, put higher value on employing younger workers (independently of whether younger workers perform differently relatively to older ones) because they represent a long-term investment or adapt more easily to a new work environment.

---

21Our model is thus quite flexible. In particular, if $s_{i,f}^{id}$ and $s_{i,f}^{co}$ are negatively correlated, we do not exclude the possibility that the more valued is a worker, the worst his future performance, a (slightly implausible) feature that was not considered in our informal discussions in section 2 concerning the direction of the bias for the estimation of $\tau$.

22Setting the payoff of losing bidders at zero is not an innocuous normalization: it means that we exclude allocative externalities à la Jehiel and Moldovanu (1996) where losing bidders care about the identity of the winning bidder. From a technical perspective, this would open the door to equilibrium multiplicity. We also exclude financial externalities à la Maasland and Onderstal (2007) where losing bidders care about the price paid by their competitors (e.g. to exhaust their budget).
In expression (2), \( \lambda \) is a strictly positive parameter, and measures to what extent performance matters in valuing workers. Note that since the valuation \( V_{i,f} \) depends on the signals \((s_{i,f}^{id}, s_{i}^{co})\) and all variables entering \( y_{i,f} \) (besides the performance shock \( \epsilon_{i,f} \)), i.e., \((s_{i,f}^{PV}, s_{i}^{CV}, w_{i}, x_{i,f})\), the payoff in (2) can be interpreted as the gain of firm \( f \) once the worker auction has ended at wage \( w_{i} \) and if the entire vector of private signals (with respect to this worker) is known to this firm, but before the worker's performance is observed (reflecting why the expectation is with respect to \( \epsilon_{i,f} \)). Note also that since the wage is allowed to influence performance, the valuation itself depends on wage as well, and hence firms' preferences do no longer satisfy the “quasi-linearity assumption” that prevails in standard auction models. If in addition \( \tau > 0 \), our labor auction model implies that firms do not necessarily wish to pay the smallest possible wage to a worker.

Let us now introduce the following renormalization of the signals: \( bs_{f}^{k} := \frac{\lambda}{1-\lambda} \cdot s_{f}^{k} \) for \( k = co, CV \) and \( bs_{f}^{k} := \frac{\lambda}{1-\lambda} \cdot s_{f}^{k} \) for \( k = id, PV \). These normalized signals will be called the “bid-signals” (hence the notation \( bs_{f} \)). They are the relevant signals for the analysis of the bidding stage. We furthermore let \( bs_{f} \) be the sum of the bid-signals observed by bidder \( f \): \( bs_{f} = bs_{f}^{id} + bs_{f}^{PV} + bs_{f}^{co} + bs_{f}^{CV} \) and \( bs_{f} = bs_{f}^{id} + bs_{f}^{PV} \) for \( f \neq inc \). The CDF of \( bs_{f}^{id} \) is denoted \( G_{f}^{id} \). We also let \( G_{f}^{PV} \) (resp. \( G_{f}^{CV} \)) denote the CDF of \( bs_{f}^{id} + bs_{f}^{PV} \) (resp. \( bs_{f}^{co} + bs_{f}^{CV} \)). All these CDFs are assumed to be atomless and, to simplify the exposition, have full support on \( \mathbb{R} \). The associated density functions are denoted by the lowercase letter \( g_{f} \), e.g., \( g_{f}^{PV} \) corresponds to the PDF of \( bs_{f}^{id} + bs_{f}^{PV} \). As detailed below, thanks to Assumption A1 ii), the structure of bidders' preferences implies that the equilibrium strategy of bidder \( f \) in the auction for worker \( i \) depends on the multi-dimensional signal \( s_{i,f} \) only through the single-dimensional aggregate bid-signal \( bs_{f} \).

For any worker \( i \) and wage \( w > 0 \), and conditional on the full set of signals of firm \( f \)'s competitors, there is a unique cutoff bid-signal in \( \mathbb{R} \) that makes firm \( f \) indifferent between hiring worker \( i \) or not. Let us characterize this cutoff first for the incumbent and then for the non-incumbents. For the incumbent it is denoted \( \tilde{bs}_{i,inc}^{id}(w) \). At this value the incumbent’s payoff equals zero, and below (resp. above) the bid-signal cutoff the payoff is positive (resp. negative). Defining \( A_{i,f}: = \frac{1}{1-\lambda} \cdot \log(\tilde{V}_{f}(x_{i,f}, z_{i,f})) + \frac{\lambda}{1-\lambda} \cdot \beta_{f} + \frac{1}{1-\lambda} \cdot \log[\mathbb{E}[e^{\lambda \cdot \epsilon_{i,f}}]] + \frac{\lambda}{1-\lambda} \cdot \beta_{f} \cdot x_{i,f} \), we have from (2) that
\[
\tilde{bs}_{i,inc}^{id}(w) = \log(w) - A_{i,inc}.
\]

Note that this cutoff does not depend on the signals of the incumbent’s competitors. The function \( w \rightarrow \tilde{bs}_{f}^{id}(w) \) is an increasing bijection from \((0, +\infty)\) to \( \mathbb{R} \). We let then \( \tilde{bs}_{f}^{id}([.) \) be the associated inverse functions.

Since non-incumbents do not observe \( bs_{f}^{co} \) and \( bs_{f}^{CV} \), their cutoff bid-signals depend on the belief they have regarding these two bid-signals. More precisely, given the payoff (2), what matters is their belief on the sum \( bs_{f}^{co} + bs_{f}^{CV} \). Given A1 ii), the distribution of \( bs_{f}^{co} + bs_{f}^{CV} \) conditional of the vector of aggregate bid-signals \( bs_{i,f}, f = 1, \cdots, F \) depends solely on \( bs_{i,inc} \).
With this in mind and letting $\theta := \lambda \cdot \tau$, we can now define the cutoff bid-signals of non-incumbents by introducing the function $U_f : \mathbb{R} \to \mathbb{R}$ given by

$$U_f(x) = \frac{1}{1-\theta} \cdot \log(E[e^{(1-\theta)(bs^{co} + bs^{CV})}|bs_f^{id} + bs_f^{PV} + bs^{co} + bs^{CV} = x]).$$ (4)

When firm $f \neq f^{inc}$ believes that the incumbent's aggregate bid-signal is $x$ (and independently of its beliefs about the other competitors' bid-signals), there is, for any $w$, a unique cutoff bid-signal $\tilde{bs}_{i,f}(w, x)$ such that the payoff of firm $f$ equals zero, and below (resp. above) which the payoff is positive (resp. negative). From (2) we have that

$$\tilde{bs}_{i,f}(w, x) = \log(w) - A_{i,f} - U_{f^{inc}}(x).$$ (5)

Compared to the incumbent, the cutoff signal value of each non-incumbent is shifted upwards or downward by $U_{f^{inc}}(x)$, and which is why this term can be interpreted as a “signal shifter”.

We make the following additional assumptions throughout our analysis:

**A2:** $\theta \equiv \lambda \cdot \tau < 1$.

**A3:** $0 < U'_{f^{inc}}(x) < 1$ for any $x \in \mathbb{R}$ and $i = 1, \ldots, N$.

**A4:** In the auction for worker $i$, the distribution $G_{f^{inc}}$, the function $U_{f^{inc}}(.)$, the scalar $A_{i,f^{inc}}$ and the parameter $\theta$ are assumed to be common knowledge across bidders.

We do not impose $\tau$ (or equivalently $\theta$) to be positive. The literature on reciprocity suggests, however, that we should expect $\tau$ to be positive. A2 imposes that $\theta$ is not too large in order to avoid the unrealistic setting in which bidders are prepared to pay any wage for a given worker. A3 is similar to the kind of assumptions frequently made in the literature on auctions with interdependent values (see Krishna (2002)), and implies that, for any value of $w$, the marginal effect of the incumbent’s signal $bs_{i,f^{inc}}$ on log $V_{i,f^{inc}}$ is larger than the marginal effect on log $E[V_{i,f}|bs_{i,f^{inc}}]$ for $f \neq f^{inc}$ (the expectation is with respect to $bs^{co} + bs^{CV}$ given the realization of $bs_{i,f^{inc}}$). A4 guarantees that any bidder $f$ is able to recover the functions $\tilde{bs}_{i,f}(.,.)$ and $\tilde{bs}_{i,f^{inc}}(.)$ (given that those functions depend on the covariates $\{(x_{i,f}, z_{i,f})\}_{f=1,\ldots,F}$ and the associated parameters $\beta$ only through the terms $A_{i,f}$ and $A_{i,f^{inc}}$).

### 3.3 Equilibrium analysis

Let us first consider the situation where there is no incumbent among the bidders in a given auction. The sum of bid-signals $bs^{co} + bs^{CV}$ is then still unknown but, since there is no incumbent, each bidder $f$ can only take the unconditional expectation with respect to this sum. The analogue of the signal shifter $U_f(x)$ is $\frac{1}{1-\theta} \cdot \log(E[e^{(1-\theta)(bs^{co} + bs^{CV})}])$. The cutoff bid-signal that makes firm $f$ indifferent between hiring or not worker $i$ at wage $w$, denoted by $\tilde{bs}_{i,f}(w)$,
is now given by
\[
\hat{b}_{i,f}(w) = \log(w) - A_{i,f} - \frac{1}{1 - \theta} \cdot \log(E[e^{(1-\theta) (bs^\alpha + bs^{\text{CV}})}]).
\] (6)

Note that the function \( \hat{b}_{i,f}(.) \), for \( f \neq f^\text{inc} \), is strictly increasing, and that each firm \( f \) has all required information to determine this function. The equilibrium strategy of each firm \( f \) consists in not entering the auction for worker \( i \) if \( bs_f < \hat{b}_{i,f}(W') \), and in remaining active until \( [\hat{b}_{i,f}]^{-1}(bs_f) \) otherwise. If the wage is assumed to have no effect on performance, i.e., \( \tau = 0 \) and hence \( \theta = 0 \), the model becomes a standard private value (PV) model where in the auction for worker \( i \), it is a dominant strategy for each firm \( f \) to bid/remain active until its valuation \( V_{i,f} = e^{[A_{i,f} + bs_{i,f}]} \cdot E[e^{(bs^\alpha + bs^{\text{CV}})}] \).

Next we consider the situation where there is an incumbent. For notational simplicity, the index \( i \) is suppressed in the remainder of this section, and the equilibrium is derived for a given vector \( \{A_f\}_{f=1,...,F} \). Since the incumbent's payoff only depends on his own aggregate bid-signal, the equilibrium strategy of the incumbent is always as in a PV model. If his aggregate bid-signal is below \( \hat{b}_{f^\text{inc}}(W') \), it is a dominant strategy not to participate in the auction; otherwise, for any given price \( p \geq W' \), it is a dominant strategy to remain active (resp. to exit) if his bid-signal is above (resp. below) \( \hat{b}_{f^\text{inc}}(p) \).

For the non-incumbents, the bidding incentives depend on their belief about the incumbent's aggregate bid-signal and whether the incumbent is still active in the auction. In order to determine whether a bidder \( f \neq f^\text{inc} \) prefers to enter or not the auction, and to remain active or to drop-out at a given position of the auction clock \( p \geq W' \), and this given the bidding history up to \( p \), we distinguish three cases.

**Case A: the incumbent has decided not to participate in the auction.** This indicates that the incumbent's aggregate bid-signal \( bs_{f^\text{inc}} \) is below \( \hat{b}_{f^\text{inc}}(W') \). From Bayesian updating, the non-incumbents can then infer that \( bs_{f^\text{inc}} \) is distributed on \((-\infty, \hat{b}_{f^\text{inc}}(W'))\] according to the distribution \( x \rightarrow G_{f^\text{inc}}(x) / G_{f^\text{inc}}(\hat{b}_{f^\text{inc}}(W')). \) Then for any \( p \geq W' \), we let \( \hat{b}^A_f(p) := \log(p) - A_f - U^A_{f^\text{inc}} \) where \( U^A_{f^\text{inc}} = \frac{1}{1 - \theta} \cdot \log(E[e^{(1-\theta) (bs^\alpha + bs^{\text{CV}})} | bs_{f^\text{inc}} \leq \hat{b}_{f^\text{inc}}(W')]) = \frac{1}{1 - \theta} \cdot \log(\int_{-\infty}^{\hat{b}_{f^\text{inc}}(W')} e^{(1-\theta)U_{f^\text{inc}}(x)} G_{f^\text{inc}}(bs^\alpha + bs^{\text{CV}})) \right) G_{f^\text{inc}}(bs^\alpha + bs^{\text{CV}})) \). \( \hat{b}^A_f(p) \) corresponds to the value of the aggregate bid-signal that makes bidder \( f \) indifferent between winning the auction or not at price \( p \), given that the incumbent has not participated in the auction.

**Case B: the incumbent has entered the auction and dropped out at price \( p^* \in [W', p) \).** Non-incumbents can then infer that the incumbent's aggregate bid-signal is exactly equal to \( \hat{b}_{f^\text{inc}}(p^*) \). We define \( \hat{b}^B_f(p, p^*) := \hat{b}_f(p, \hat{b}_{f^\text{inc}}(p^*)) \), which is the value of the aggregate bid-signal that makes bidder \( f \) indifferent between winning the auction or not at \( p \), given that the

\[\text{We can rewrite}\frac{1}{1 - \theta} \cdot \log(E[e^{(1-\theta)(bs^\alpha + bs^{\text{CV}})}]) = \frac{1}{1 - \theta} \cdot \log(\int_{-\infty}^{\infty} e^{(1-\theta)U_{f^\text{inc}}(x)} G_{f^\text{inc}}(x)) \text{ for any possible } f^\text{inc}.\]

Given A4, \( \theta \), \( G_{f^\text{inc}}(.) \), and \( U_{f^\text{inc}}(.) \) (defined in (4)) are known, so (6) can indeed be determined by each firm \( f \).
incumbent has dropped out when the clock reached $p^*$.

**Case C: the incumbent has entered the auction and is still active at price** $p$. We let $\hat{bs}_f^C(p) := \hat{bs}_f(p, \hat{bs}_{f,inc}(p))$, which corresponds to the aggregate bid-signal that makes bidder $f$ indifferent between winning the auction or not at $p$, as if he knew that the incumbent would instantly exit the auction at $p$, which would reveal that the incumbent’s aggregate bid-signal is equal to $\hat{bs}_{f,inc}(p)$. Thanks to A3, $\hat{bs}_f^C(p)$ is strictly increasing in $p$ on $[W^r, +\infty)$.  

The equilibrium corresponding to the situation where an incumbent is present is given in Proposition 3.1.

**Proposition 3.1.** In equilibrium, the strategy of the incumbent consists in not entering the auction if $bs_{f,inc}$ is below $\hat{bs}_{f,inc}(W^r)$, and in remaining active until $[\hat{bs}_{f,inc}]^{-1}(bs_{f,inc})$ otherwise. The strategy of non-incumbents depends on the information revealed about the incumbent’s bidding behavior:

- **Participation decision of non-incumbents:** If the incumbent has not entered the auction, then the non-incumbent firm $f$ should enter only if $bs_f \geq \hat{bs}_f^A(W^r)$. If the incumbent has entered, then $f$ should enter only if $bs_f \geq \hat{bs}_f(W^r, \hat{bs}_{f,inc}(W^r))$.

- **Dropout decision of non-incumbents:** Suppose the non-incumbent firm $f$ has entered the auction, and the auction clock has reached $p$. If the incumbent has not entered (resp. dropped out at $p^* < p$), then $f$ should exit instantly if $bs_f < \hat{bs}_f^A(p)$ (resp. $bs_f < \hat{bs}_f^B(p, p^*)$), and remain active otherwise. If the incumbent is still active at $p$, then $f$ should exit instantly if $bs_f < \hat{bs}_f^C(p)$, and remain active otherwise.

The “inverse bidding functions” $\hat{bs}_{f,inc}^*(\cdot)$, $\hat{bs}_f^A(\cdot)$, $\hat{bs}_f^B(\cdot; p^*)$ (for any $p^* \geq W^r$) and $\hat{bs}_f^C(\cdot)$ are increasing. Given A4, each bidder $f$ disposes of all necessary information to compute these functions.

The equilibrium strategy of each non-incumbent depends on whether the incumbent is active or not. If the incumbent is not active at $p$, all non-incumbents have the same beliefs about the incumbent’s signal. From a strategic perspective, their bidding incentives are similar to the ones in a PV model: it is a dominant strategy for non-incumbents to remain active at $p$ if and only if their aggregate bid-signal is above the cutoff value signal that makes them indifferent between winning the auction or not at $p$. If, however, the incumbent is still active at $p$, our auction model is fundamentally an interdependent value model. When deciding whether or not to remain active, bidders should in particular anticipate that the auction state can switch from case C to case B and the corresponding continuation value. The resolution is similar to Milgrom and Weber (1982): in order to avoid the winner’s curse, non-incumbents should behave as if the incumbent is going to quit instantly the auction. They should thus act as if they are in the

\[24\text{Formally, this is because } \frac{d\hat{bs}_f^C(p)}{dp} = \frac{1}{p} \cdot [1 - U'_{f,inc}(\hat{bs}_{f,inc}(p))] > 0.\]
worst-scenario wherein the incumbent has the lowest possible signal, explaining why \( f \) should instantly quit if \( bs_f \) is strictly below \( \hat{bs}_C(p) \), and stay active otherwise. The formal proof can be found in the Appendix.

Before ending this section we wish to make several remarks.

1) Our equilibrium analysis extends easily if the term \( \tau \cdot \log(w) \) is replaced by some non-specified function \( \rho(w) \). Assumption A2 should then be replaced by two restrictions on this function: \( \rho'(w) < \frac{1}{\lambda w} \) and \( \lim_{w \to +\infty} w \cdot e^{-\lambda \rho(w)} = +\infty \). Under these restrictions it can be shown that the various inverse bidding functions \( \hat{bs} \) remain strictly increasing. 2) It is a priori not excluded that a non-incumbent prefers to drop-out immediately after seeing the incumbent dropping out, which would be a source of bidding ties. However, given the equilibrium bidding strategies, this event would never occur with positive probability on the equilibrium path.\textsuperscript{25} 3) The main take-away from this section, and which will be used in our econometric analysis, is that the equilibrium is unique and that equilibrium strategies are increasing in the aggregate bid-signals. Those are the key properties that need to be derived if one wants to adapt easily our approach to other auction mechanisms. Note that such properties are verified in a large class of standard auctions (see Athey (2001)).

4 The econometric methodology

This section is devoted to the econometric aspects of the paper. Section 4.1 establishes conditions under which the primitives of the auction model and the parameters of the performance equation are identified (further discussions on identification are gathered in Section 5). In Section 4.2 and its associated Appendix, we show how to compute in general the two terms that we plug into the performance equation in order to control for the various sources of endogeneity. Furthermore, we provide details on the expressions obtained under a Gaussian structure. Section 4.3 presents briefly the estimation method.

4.1 Identification

Identification is obtained in two steps: 1) From the bidding data we identify the distribution of the bid-signals, i.e., the CDFs \( G_f^{PV} \) for \( f = 1, \ldots, F \) and the CDF \( G^{CV} \), and also the parameter \( \theta \) and the expression of \( A_{i,f} \) as a function of the covariates \( x_{i,f} \) and \( z_{i,f} \). This is a non-parametric identification result in the sense that the bid-signal distributions are left unspecified. 2) From the first step, we can identify the various bidding functions \( \hat{bs} \) associated to each worker which

\textsuperscript{25}Suppose the incumbent drops out at \( p^* \) while there are at least two active non-incumbents, implying that at this price we have switched from case C to case B. Let \( x \) be the private signal of a non-incumbent \( f \) that has decided to remain active until \( p^* \): we have then \( \{ \hat{bs}_C \}^{-1}(x) \geq p^* \), or equivalently \( x \geq \hat{bs}_f(p^*, \hat{bs}_{inc}(p^*)) \). It is then only in the degenerate case where \( x = \hat{bs}_f(p^*, \hat{bs}_{inc}(p^*)) \) that firm \( f \) would like to drop out immediately at \( p^* \) as well.
enables us then to identify to two control terms that we plug in the performance equation (1) to control for the various sources of endogeneity. Finally, the augmented performance equation enables us to identify the coefficients $\tau$ and $\beta_x$.

Before introducing some additional assumptions, let us introduce some additional notation. Let $S_i \subseteq \{1, \cdots, F\}$ denote the publicly observable subset of firms that are potential participants in the auction for worker $i$.\footnote{In a procurement setup, $S_i$ would correspond to the set of firms that have passed a qualification phase in order to be allowed to submit an eligible bid.} We also decompose the vector $x_{i,f}$ as $(x_{i,f}^1, x_{i,f}^2, x_{i,f}^{dum})$ where $x_{i,f}^2 = 0$ if firm $f$ is not the incumbent for $i$ (or equivalently if $x_{i,f}^{dum} = 0$). The vector $x_{i,f}^1$ corresponds then to the set of characteristics that are relevant when $f$ is not an incumbent firm. Note that the vector $x_{i,f}^2$ can include interaction terms between $x_{i,f}^1$ and $x_{i,f}^{dum}$.

A5: The information set $\mathcal{I}$ contains the variables $z_{i,f}$, $x_{i,f}$, $S_i$ and $W_i^{f}$ for all $i$ and $f$, the identity of the incumbent $f_i^{inc}$ -if any- for all workers $i$, and also $w_i$ and $f_i^{w}$ for all workers $i$ that are actually sold.

A6: The signals $(s_{i}^{co}, s_{i}^{CV})$ are drawn independently of the signals $\{(s_{i,f}^{id}, s_{i,f}^{PV})\}_{f=1,\cdots,F}$.

A7: $\log(V_f(x_{i,f}, z_{i,f})) = \alpha_f + \alpha_x \cdot x_{i,f} + \alpha_z \cdot z_{i,f}$. For any given bidder $f$ and set of potential participants $S$ with $f \in S$, i) the vector composed of 1, the characteristics $x_f$ and the auction covariates $z_f$ is of full rank on a positive measure of the observables on the subsample without incumbents and the set of potential participant is $S$; ii) the vector composed of 1 and the characteristics $x_f^2$ is of full rank on a positive measure of the observables on the subsample where firm $f$ is the incumbent and the set of potential participant is $S$.

A8: The set of potential participants $S_i$ is drawn independently of the bid-signals, and we have: i) For each $i$ there is a positive probability that $S_i$ involves an incumbent (i.e. $f_i^{inc} \in S_i$), and a positive probability that it does not involve one. ii) For any firm $f \in \{1, \cdots, F\}$, within the subsample of auctions $i$ with $f_i^{inc} \notin S_i$, $f \in S_i$ and with the covariates $(x_f, z_f)$ for firm $f$, there is a strictly positive probability that $S_i$ contains at least two bidders. iii) In the subsample of auctions $i$ with an incumbent, there is a positive probability that $S_i$ contains exactly two bidders.

A5 defines $\mathcal{I}$, the set of variables that is assumed to be observed by the econometrician. Note that it contains the explanatory variables appearing in performance equation (1) not only for the winning firm but also for losing bidders. Concerning the auction process, observing the final price and the identity of the winner corresponds here to the minimal assumption we can make.\footnote{Discrete choice models can be interpreted as a pure private value auction model. Nevertheless, from the econometrician perspective, it would be as if the auction price were not observed. If $\tau = 0$ and if only the identity of the winner is observed, our econometric problem would then reduce to the polychotomous selection problem addressed by Dubin and McFadden (1984). The observation of the auction price does not solely allow us to deal with $\tau \neq 0$ but also to identify non-parametrically the underlying signal/valuation distributions (instead of imposing extreme value distributions as in Dubin and McFadden (1984)).} A6 is a crucial independence assumption which enables us to use a deconvolution
argument: once the distributions $G_f$ and $G_{PV}^f$ are identified for a given firm $f$, then we can recover the distribution $G^{CV}$, which is a key step to identify the functions $U_{i,f|inc}(.)$ and so the bidding cutoffs associated to observed behavior. The main role of A7 is to simplify the presentation as explained later. A8 is our fundamental exclusion restriction: in particular, the bid-signal distribution is the same in the samples with and without incumbents, which enables us to disentangle the private value component from the common value component in the bidding behavior. However, we stress that thanks to the term $x_i^2$, we allow the impact of the incumbency status on valuations to be much richer than a simple additive shift. A8 ii) is a weak assumption that allows to identify $G_{PV}^f$ non-parametrically for each $f$. It could be dropped under the assumption that signal distributions are symmetric across bidders. A8 iii) is a condition which simplifies the proof and that can be relaxed with the use of covariates as discussed in Section 5.

4.1.1 Identification of the auction model

To simplify, we assume in this section that the reserve price is always zero: $W_i^r = 0$ for all $i$. We first use the subsample of auctions $i$ for which the set $S_i$ does not include an incumbent. Given A7, the cutoff bid-signal (6) can be rewritten as

$$\hat{b}_s(w_i) = \log(w_i) - \beta_f^* - \beta_x^* \cdot x_{i,f} - \beta_z^* \cdot z_{i,f} \equiv \log(w_i) - A_f^*$$

where $\beta_f^* = \frac{1}{1-\theta} \cdot \left[ \lambda \beta_f + \alpha_f + \log\left( E[e^{\lambda x_{i,f}}] \right) + \log(\int_{-\infty}^{+\infty} e^{(1-\theta) x \cdot dG^{CV}(x)}) \right]$, $\beta_x^* = \frac{1}{1-\theta} \cdot [\lambda \beta_x + \alpha_x]$ and $\beta_z^* = \frac{1}{1-\theta} \cdot \alpha_z$. Below, we use the notation $\beta_x^{1,*}, \beta_x^{2,*}$ and $\beta_x^{dum,*}$ for the coefficients associated to $(x_{i,f}^1, x_{i,f}^2)$ and $x_{i,f}^{dum}$, respectively.

From Athey and Haile (2002) we know that in the English (button) auction with no reserve price, at least two bidders and with no covariates, then the independent private value model is non-parametrically identified from the joint observation of the identity of the winner and the winning price. Applying Athey and Haile (2002) to the subsample without incumbents allows us to identify $G_{PV}^f$ and $\beta_f^*$ for each $f$ and also the vector of coefficients $(\beta_x^{1,*}, \beta_x^{2,*})$ as detailed in next paragraph.

Take a given firm $f^*$. Applying Athey and Haile (2002) on any given point $\{(x_f, z_f)\}_{f=1, \ldots, F}$ in the support of the observable covariates and for a given set of participants $S$ such that $|S| \geq 2$ and $f^* \in S$ (with arises with positive probability given A8 ii)), the (centered) distributions $G_{PV}^f$, for each $f \in S$, and the associated value for $(A_f^*)_{f \in S}$ are thus identified. This enables us to identify $G_{PV}^f$ for each $f$. Then from A7, there exists a firm $f$ and a family of associated

---

28See Footnote 7 for the use of similar exclusion restrictions in the literature.

29We discuss in Section 5 how binding reserve prices would help if we are prepared to assume that reserve prices are fixed exogenously.
covariates \((x_f[k], z_f[k])_{k=1,...,K}\) so that \((1, x_f^1[k], z_f[k])\) is of full rank and such that those realizations of covariates belong to the support of the observables among the subsample of auctions solely made up of non-incumbents. From the corresponding vector \(\{A_f^*[k]\}_{k=1,...,K}\) that is identified, we recover the vector of coefficients \((\beta_f^*, \beta_x^{1,*}, \beta_z^*)\).

In words, the rest of our identification argument works roughly as follows: 1) the comparison of the bid-distribution of a bidder in an environment without incumbent with the bid distribution when being the incumbent allows to identify the distribution \(G^{CV}\); 2) the comparison of the bid-distribution of a bidder in an environment without incumbent with the bid distribution when facing an incumbent allows to identify the parameter \(\theta\).

Next we use the subsample of auctions with an incumbent to identify the distribution \(G^{CV}\), the parameter \(\theta\) and also \(\beta_x^{dum,*}\) and \(\beta_x^{2,*}\). More precisely, we use the subsample of auctions \(i\) with a given incumbent \(f^{inc}\) and a given non-incumbent \(f^*\) and such that \(S_i = \{f^{inc}, f^*\}\). We restrict ourselves to auctions with two bidders for the following reason: with at least three bidders, we could no longer apply Athey and Haile (2002) to recover the distribution of the drop-out price of each bidder, and this because drop-out prices would not be independently distributed because those bids are not fixed ex-ante as a function of the private signal of the corresponding bidder but would depend on the bidding history as characterized in Proposition 3.1. By contrast, with two bidders, we have the following two properties: 1) the drop-out price of each bidder does not vary across the bidding history since the auction stops immediately after the first drop-out; 2) the drop-out prices of the two bidders are distributed independently since their private signals are distributed independently.

Using the aforementioned auctions with two bidders, we can identify the distributions of the price at which bidders \(f^{inc}\) and \(f^*\) quit the auction. For any given realization of the covariates on the support of the observables, we then identify the CDF \(G_{f^{inc}}\) and the scalar \(A_{f^{inc}}\) from the distribution of the drop-out price of the incumbent. Since \(G_{f^{inc}}^{PV}\) has been identified from the subsample without incumbents (for each \(f\) and in particular for \(f^{inc}\)) and given A8, a standard deconvolution argument (see, e.g., Diggle and Hall (1993)) now implies, thanks to A6, that \(G^{CV}\) is identified as well. The general expression of \(G^{CV}\) as an explicit function of \(G_{f^{inc}}^{PV}\) and \(G_{f^{inc}}\) is given in Appendix B.1.\(^{30}\)

The identification of \(A_{f^{inc}}\) for a set of covariates such that the associated vector \((1, x_{f^{inc}}^2)\) is of full rank (A7), enables us to identify the vector of coefficients \(\beta_x^{2,*}\) and the constant

\[
\beta_x^{dum,*} + \frac{1}{1-\theta} \left[ \lambda \beta_{f^{inc}} + \alpha_{f^{inc}} + \log(E[e^{\lambda x_{f^{inc}}^{inc}}]) \right] = \beta_x^{dum,*} + \beta_{f^{inc}}^* - \frac{1}{1-\theta} \log \left( \int_{-\infty}^{+\infty} e^{(1-\theta)x} dG^{CV}(x) \right)
\]

(7)

where \(\beta_{f^{inc}}^*\) has been identified from the bidding data without incumbent and where \(G^{CV}\) has

\(^{30}\)See Li, Perrigne, and Vuong (2000) for a seminal application of Fourier transformations to the structural econometrics of auctions and in particular to develop nonparametric estimators. By contrast, the estimators we develop are relying on a Gaussian structure, which simplifies considerably the deconvolution formula for \(G^{CV}\).
also been identified. (7) implies that if \( \theta \) is identified, then \( \beta_{x}^{\text{dum},*} \) will be identified as well.

Let us now show that the parameter \( \theta \) is identified thanks to the distribution of the drop-out price of firm \( f^* \) in the sample where \( f^* \) faces incumbent \( f^{\text{inc}} \) and for a given set of covariates yielding the bid shifters \( A_{f} \), and \( A_{f^{\text{inc}}} \). Let \( G^{\text{CV}}(\cdot|\bar{x}, f^{\text{inc}}) \) denote the CDF of \( b_{s}^{\text{CV}} \) conditional on \( b_{s}^{f^{\text{inc}}} = \bar{x} \). Under A6, the associated PDF, denoted by \( g^{\text{CV}}(\cdot|\bar{x}, f^{\text{inc}}) \), satisfies \( g^{\text{CV}}(x|\bar{x}, f^{\text{inc}}) = \frac{g^{\text{CV}}(x,g_{s}^{\text{inc}}(\bar{x}-x))}{g_{s}^{\text{inc}}(\bar{x})} \), which is known since all densities on the right hand side have already been identified. Using this additional notation, (4) can be written as

\[
U_{f^{\text{inc}}}(\bar{x}) = \frac{1}{1-\theta} \cdot \log \left[ \int_{-\infty}^{+\infty} e^{(1-\theta)x} \cdot dG^{\text{CV}}(x|\bar{x}, f^{\text{inc}}) \right].
\]

We get from (5) that the drop-out price \( p \) of the non-incumbent \( f^* \) with the bid-signal \( b_{s}^{f^*} \), is characterized by the equation

\[
b_{s}^{f^*} = \log(p) - A_{f^*} - U_{f^{\text{inc}}}(\log(p) - A_{f^{\text{inc}}}).
\]

Conditional on the scalar \( \theta \), the equilibrium drop-out price of the non-incumbent \( f^* \) as a function of the sum of its bid-signals \( b_{s}^{f^*} \) can thus be expressed as \( \Lambda(b_{s}^{f^*}, \theta) \), where the function \( \Lambda \) has been identified from the previous steps.

We then use the following result (see Proposition 2 in Gollier (2001)):

**Lemma 1.** If \( Y \) is a stochastic variable with \( \text{Var}(Y) \neq 0 \), then the function \( z \to \frac{1}{z} \log(E[e^{zY}]) \) is (strictly) increasing on \((0, +\infty)\).

Given the expression of \( U_{f^{\text{inc}}} \) and Lemma 1, we get that \( \Lambda(b_{s}^{f}, \theta) \) is increasing in \( \theta \). Take a given quantile \( q \in (0,1) \) of the bid-distribution of bidder \( f^* \), denoted by \( p^{q} \). Since the bid function is increasing, it corresponds then to the bid associated to the quantile \( q \) of the signal-distribution \( G_{f}^{\text{PV}} \), denoted by \( b_{s}^{q} \). We have shown above that the drop-out price distribution of \( f^* \) has been identified and thus \( p^{q} \), while \( b_{s}^{q} \) has been identified from the bidding data without incumbents. At most one value of \( \theta \) can then be consistent with \( p^{q} = \Lambda(b_{s}^{q}, \theta) \) which proofs that \( \theta \) is identified.

Our identification result regarding the primitives of the auction model is summarized in the following proposition.

**Proposition 4.1.** In the English auction without reserve prices, the distributions \( G^{\text{CV}} \) and \( G_{f}^{\text{PV}} \), \( f = 1, \cdots, F \), the scalars \( \theta \) and \( \beta_{f}^{*}, f = 1, \cdots, F \), and the vectors \( \beta_{x}^{*} \) and \( \beta_{z}^{*} \) are identified from the bidding data under A1-A8.

From Proposition 4.1, we have identified all the primitives that are necessary to compute equilibrium strategies as a function of the bid-signals, which are in turn essential to determine the precise form of the two terms we use as control functions (see Section 4.2).

However, we should be cautious that if we were in a private value environment (if \( b_{s}^{\text{CV}} \equiv 0 \), so that \( G^{\text{CV}}(\cdot) \) would be a Dirac distribution), then our identification argument to identify \( \theta \) would no longer work (formally we could no longer apply Lemma 1 since \( \text{Var}(b_{s}^{\text{CV}}) = 0 \). Still, the impossibility to identify \( \theta \) does not prevent us to identify the coefficient \( \beta_{x}^{\text{dum},*} \) through eq. \[31\]

Assumption A3 guarantees that this equation has a unique solution.
(7) where the last term would actually shrink to null. On the whole, under private values, we would identify the distributions \( G_f^{PV}, f = 1, \cdots, F \), the scalars \( \beta^*_1, f = 1, \cdots, F \), and the vectors \( \beta^*_x \) and \( \beta^*_z \) which would still enable us to compute equilibrium strategies as a function of the bid-signals<sup>32</sup> and then to pursue our control function approach.

Before ending this section we wish to make a few extra comments. 1) To simplify the presentation, we have assumed that the vectors \( \beta^*_x \) and \( \beta^*_z \) do not depend on \( f \). However, we could easily deal with such asymmetries (but then A8 iii) should be strengthened a bit). 2) The set of covariates \( x_{i,f} \) could possibly include the dummy variable \( \sum_{f=1}^{F} x_{i,f}^{dum} \) which captures whether the worker was employed in one of the firms. More generally, we allow some forms of asymmetries between the samples with and without incumbents (we could also include interaction terms between the dummy \( \sum_{f=1}^{F} x_{i,f}^{dum} \) and some characteristics of the worker). 3) To show identification, we have used only a given quantile of the drop-out price distribution of a given non-incumbent in the sample with an incumbent. Contrary to the independent pure private value model under A5, our model is thus over-identified. 4) The bidding data alone do not allow us to identify the key parameters, \( \beta \) and \( \theta \) (\( = \lambda \cdot \tau \)). However, the performance data are needed to get the point estimates of \( \tau \) and \( \beta_x \) as developed next. 5) As Athey and Haile (2008), we could relax A7 and allow for a more flexible (e.g. non-parametric) specification of \( \nabla_f (x_{i,f}, z_{i,f}) \) and then of \( A_f \) as a function of \( x_{i,f} \) and \( z_{i,f} \).

### 4.1.2 Identification of the performance equation

The control function approach amounts to modeling the conditional expectation \( E(u_{i,f}^{PV} | \mathcal{G}) \) using our auction model and bidding data. By definition of the bid-signals we have

\[
E(u_{i,f}^{PV} | \mathcal{G}) = \frac{1 - \lambda \tau}{\lambda} \cdot E[b_{i,f}^{PV} + b_{i}^{CV} | \mathcal{G}].
\]  

(8)

The following assumption imposes some parametric restrictions linking the distributions of \( b_{i,f}^{id} \) and \( b_{i,f}^{PV} \) for each \( f \), and also the the distributions of \( b_{i}^{co} \) and \( b_{i}^{CV} \).

\textbf{A9}: \( E[b_{i,f}^{PV} | b_{i,f}^{id} + b_{i,f}^{PV} = x] = \sum_{l=1}^{L} d_{i,f}^{PV} \cdot x^l \) and \( E[b_{i}^{CV} | b_{i}^{co} + b_{i}^{CV} = x] = \sum_{l=1}^{L} d_{i}^{CV} \cdot x^l \), for any \( x \) and each \( f \).

We do not impose any restriction on \( L \) leaving thus much flexibility.<sup>33</sup> Nevertheless, to obtain simple expressions for our control terms, we will consider the case \( L = 1 \) in Section and later in our simulations and application. With the extra restriction, A9 is referred to as A9<sup>9</sup>.

<sup>32</sup>Under private values, the re-normalization of signals into bid-signals allow us precisely to get rid of the dependence of \( \theta \) in the equilibrium strategies.

<sup>33</sup>Note that we could also use an alternative basis of function instead of the polynomial basis \( x, x^2, \cdots, x^L \).
For each \( l = 1, \ldots, L \), let us introduce the notation:

\[
C_{i,f}^{PV}[l] = E[(b_{i,f}^{id} + b_{i,f}^{PV})|\mathcal{G}] \quad \text{and} \quad C_{i}^{CV}[l] = E[(b_{i}^{co} + b_{i}^{CV})|\mathcal{G}].
\]

and \( \gamma_{l,f}^{PV} = d_{l,f}^{PV} \cdot \frac{1-\lambda}{\lambda} \) and \( \gamma_{l}^{CV} = d_{l}^{CV} \cdot \frac{1-\lambda}{\lambda} \).

From the law of iterated expectations, we have \( E[b_{s,i,f}^{PV}|\mathcal{G}] = E[E[b_{s,i,f}^{PV}|\mathcal{G}, b_{s,i,f}^{id}, b_{s,i,f}^{PV}]|\mathcal{G}] \) and \( E[b_{s,i}^{CV}|\mathcal{G}] = E[E[b_{s,i}^{CV}|\mathcal{G}, b_{s,i}^{co} + b_{s,i}^{CV}]|\mathcal{G}] \). Given the independence assumptions \( A1 \) and \( A6 \), and the fact that the bidding strategy of each firm \( \xi \), the relevance of this “full rank condition” will be discussed in Section 4.2. Under private values, these terms are by construction uncorrelated to all regressors appearing in (10).

The terms \( C_{i,f}^{PV}[l] \) and \( C_{i}^{CV}[l] \), for any \( l = 1, \ldots, L \), are identified: from the bidding data, we have identified the distributions of the bid-signals \( b_{s,i,f}^{id} + b_{s,i,f}^{PV} \) and \( b_{s,i}^{co} + b_{s,i}^{CV} \) and also the function that maps those signals into the observable bidding history (thanks to Proposition 3.1). Computing the terms \( C_{i,f}^{PV}[l] \) and \( C_{i}^{CV}[l] \), for any \( l = 1, \ldots, L \), reduces thus to a Bayesian updating exercise. Nevertheless, this exercise is slightly tedious in the English auction with interdependent values, and is thus detailed a bit in Section 4.2.

We obtain then the augmented performance equation

\[
y_{i,f} = \beta_{f} + \beta_{x} \cdot x_{i,f} + \tau \cdot \log(w_{i}) + \sum_{l=1}^{L} \gamma_{l,f}^{PV} \cdot C_{i,f}^{PV}[l] + \sum_{l=1}^{L} \gamma_{l}^{CV} \cdot C_{i}^{CV}[l] + \xi_{i,f}.
\]

where the error term \( \xi_{i,f} = u_{i,f} - E[u_{i,f}|\mathcal{G}] \) is by construction uncorrelated to all regressors appearing in (10).\(^{34}\)

If the vector composed of \( 1, x_{i,f}, \log(w_{i}), C_{i,f}^{PV}[l] \) and \( C_{i}^{CV}[l] \), \( l = 1, \ldots, L \) is of full rank, then the augmented performance equation (10) allows us to identify the coefficients \( \beta_{f}, \gamma_{l,f}^{PV} \) and \( \gamma_{l}^{CV} \) \( f = 1, \ldots, F \) and \( l = 1, \ldots, L \), \( \tau \) and the vector \( \beta_{x} \) thanks to the performance data. The relevance of this “full rank condition” will be discussed in Section 4.2. Under private values, note that \( C_{i}^{CV}[l] = 0 \) for each \( l \) and that the full rank condition will not be satisfied.

Finally, if we combine what is identified from the bidding and performance data, we can go

\(^{34}\text{There is no correlation because \( \mathcal{G} \) includes all explanatory variables in (10), those appearing directly in this model \( x_{i,f} \) and \( w_{i} \), and those appearing indirectly through the control terms \( C_{i,f}^{PV}[l] \) and \( C_{i}^{CV}[l] \).}
further if $\tau \neq 0$: The parameter $\lambda$ is identified since $\lambda = \frac{\theta}{\tau}$. This in turn implies that $\alpha_x$ is identified through the relationship $\alpha_x = (1-\theta) \cdot \beta_x^* - \lambda \cdot \beta_x$.

Under private values, the regression (10) is no longer appropriate and should be replaced by

$$y_{i,fw} = \beta_{fw} + \beta_x \cdot x_{i,fw} + \tau \cdot \log(w_i) + \sum_{l=1}^{L} CF_{iw}^{PV} [1] + \xi_{i,fw}.$$  

Given that $\theta$ is not identified from the first stage, it implies that we are not able to identify $\lambda$ under private values and this even if $\tau \neq 0$. Formally, two different pairs $(\lambda, \theta)$ and $(\lambda', \theta')$ will lead to the same bid and performance outcomes if $\frac{\lambda}{1-\theta} = \frac{\lambda'}{1-\theta'}$. In practice if we are close to a pure private value model, this suggests that some parameters of the model (like the coefficients $\theta$, $\lambda$, $d_{CV}^f$, and $d_{PV}^f$) will be poorly estimated. This is consistent with our simulation study (Section 5), but, fortunately, the estimates of the key parameters $\beta_x$ and $\tau$ remain satisfactory.

4.2 Control functions

In this section we detail how the control functions $CF_{iw}^{PV} [1]$ and $CF_{iw}^{CV} [1]$ are characterized in general. The precise expressions of these functions require Bayesian updating conditional on the information set $\mathcal{I}$ given the equilibrium behavior from Proposition 3.1. We also illustrate our computations for $CF_{iw}^{PV} [1]$ and $CF_{iw}^{CV} [1]$ assuming that bid-signals follow a symmetric Gaussian structure as defined next.

**Definition 1.** Bid-signals follow a symmetric Gaussian structure if the bid-signals $bs_{id}^{fw} + bs_{PV}^{fw}$, $f = 1, \cdots, F$ and $bs_{CV}^{co} + bs_{CV}^{f}$ are distributed independently in the following way: i) $bs_{id}^{fw} + bs_{PV}^{fw}$ has the same distribution for each $f$ and the corresponding distribution $G_{PV}^f$ is a centered normal distribution with variance $\sigma_{PV}^2$, ii) $G_{CV}^f$ is a centered normal distribution with variance $\sigma_{CV}^2$.

Letting $\Phi$ denote the CDF of a standard normal distribution, under a symmetric Gaussian structure we have $G_{k}^f(x) = \Phi(\frac{x}{\sigma_k^f})$ for $k = PV, CV$. Under these normality assumptions and for $L = 1$ (coupled with our other assumptions), the two control functions have explicit and tractable expressions. These expressions are derived using a series of well known properties on (truncated) normally distributed variables (which can be found in for example [ ]). Under the symmetric Gaussian structure, we also note that A3 is always satisfied (see 23 in Appendix).

The precise form of the control functions depends on whether an incumbent is present among the potential auction participants, and, if there is an incumbent, on the identity of the auction winner. We distinguish thus four cases: 1) There is no incumbent; 2) The winner is the incumbent; 3) The winner is not the incumbent and the auction clock stops at the reserve price (so that there is no second-highest bidder with probability one); 4) The winner is not the incumbent and the auction price is strictly above the reserve price (so that there is a second-highest bidder who can be either the incumbent or a non-incumbent).

**Case 1:** In the absence of an incumbent, worker $i$ is sold to firm $f_{iw}$ at wage $w_i \geq W_i^r$ if
and only if $bs_{i,fw} \geq \bar{bs}_{i,fw}(w_i)$ and $\max\{\max_{f \neq f'}[[\bar{bs}_{i,f}^{-1}(bs_{i,f})], W'_f\} = w_i$. Recall that the function $\bar{bs}_{i,f}(p)$ is given by (6) for $f \neq f'$. Given the independence assumptions in A1, we have that conditional on $\mathcal{S}$, $bs_{i,fw}$ is distributed according to the distribution $G_{f,inc}^{PV}$ truncated below $\bar{bs}_{i,fw}(w_i)$, i.e., it has the distribution function $\frac{G_{f,inc}^{PV}(\bar{bs}_{i,fw}(w_i)) - G_{f,inc}^{PV}(\bar{bs}_{i,fw}(w_i))}{1 - G_{f,inc}^{PV}(\bar{bs}_{i,fw}(w_i))}$ on the interval $[\bar{bs}_{i,fw}(w), +\infty)$. Since there is no incumbent in this case, the independence assumptions A1 and A6 guarantee that there is no updating on $bs_{i}^{co} + bs_{i}^{CV}$. We therefore have

$$CF_{i,fw}[l] = \int_{\bar{bs}_{i,fw}(w_i)}^{+\infty} x^l \cdot \frac{d[G_{f,inc}^{PV}(x)]}{1 - G_{f,inc}^{PV}(\bar{bs}_{i,fw}(w_i))}$$

$$CF_{i,fw}^{PV}[l] = 0.$$

Under the symmetric Gaussian structure and for $l = 1$, (11) reduces to

$$CF_{i,fw}^{PV}[1] = \sigma_{PV} \cdot \frac{\phi\left(\frac{\bar{bs}_{i,fw}(w_i)}{\sigma_{PV}}\right)}{1 - \phi\left(\frac{\bar{bs}_{i,fw}(w_i)}{\sigma_{PV}}\right)} = \sigma_{PV} \cdot \frac{\phi\left(\frac{\log(w_i) - A_{\bar{bs}_{i,fw}}}{\sigma_{PV}}\right)}{1 - \phi\left(\frac{\log(w_i) - A_{\bar{bs}_{i,fw}}}{\sigma_{PV}}\right)},$$

where $\phi(.)$ is the PDF of a standard normal distribution. The control function $CF_{i,fw}[1]$ corresponds then to an inverse Mills ratio multiplied by the standard deviation $\sigma_{PV}$.

In (12), $CF_{i,fw}^{PV}[1]$ is a non-linear function of $\log(w_i)$, and the variables $x_{i,fw}^1$ and $z_{i,fw}$. If we limit ourselves to the sample without incumbents then the variables $1, x_{i,fw}^1, \log(w_i)$ and $CF_{i,fw}^{PV}[1]$ will be of full rank for each $f$, which implies that the associated coefficients $\beta_f$, $\beta_f^1$, $\tau$, and $\gamma_f^{PV}$ are identified. Naturally, the parameter $\gamma^{CV}$, however, is not identifiable since $CF_{i}^{CV}$ does not vary in the sample containing only non-incumbents.\(^{35}\)

**Case 2:** From Proposition 3.1 we know that the incumbent is the winner ($f'_i = f'^{inc}_i$) if and only if $bs_{i,f'^{inc}_i} \geq \bar{bs}_{i,f'^{inc}_i}(w_i)$ and $\max\{\max_{f \neq f'^{inc}_i}[[\bar{bs}_{i,f}^{-1}(bs_{i,f})], W'_f\} = w_i$, with $w_i \geq W'_f$. Given the independence assumptions in A1, $bs_{i,f'^{inc}_i}$ conditional on $\mathcal{S}$ is distributed according to the distribution $G_{i,f'^{inc}_i}$ truncated below $\bar{bs}_{i,f'^{inc}_i}(w_i)$, i.e., it has the distribution function $\frac{G_{i,f'^{inc}_i}(\bar{bs}_{i,f'^{inc}_i}(w_i)) - G_{i,f'^{inc}_i}(\bar{bs}_{i,f'^{inc}_i}(w_i))}{1 - G_{i,f'^{inc}_i}(\bar{bs}_{i,f'^{inc}_i}(w_i))}$ on the interval $[\bar{bs}_{i,f'^{inc}_i}(w), +\infty)$. Furthermore, the distribution of $bs_{i}^{co} + bs_{i}^{CV}$, conditional on $\mathcal{S}$ and $bs_{i,f'^{inc}_i}$, only depends on the aggregate bid-signal of the incumbent (i.e., $\mathcal{S}$ drops out). Noting that $bs_{i,f'^{inc}_i}^{inc} + bs_{i,f'^{inc}_i}^{PV} = bs_{i,f'^{inc}_i}^{inc} - bs_{i}^{co} + bs_{i}^{CV}$, we therefore have

$$CF_{i,f'^{inc}_i}[l] = \int_{\bar{bs}_{i,f'^{inc}_i}(w_i)}^{+\infty} E_{x \sim G^{CV}(\bar{bs}_{i,f'^{inc}_i})}[|x - x'|] \cdot \frac{dG_{i,f'^{inc}_i}(\bar{bs}_{i,f'^{inc}_i}(w_i))}{1 - G_{i,f'^{inc}_i}(\bar{bs}_{i,f'^{inc}_i}(w_i))} = \int_{\bar{bs}_{i,f'^{inc}_i}(w_i)}^{+\infty} \left[ \int_{-\infty}^{x} (x - x') \cdot g^{CV}(x)g^{PV}_{i,f'^{inc}_i}(x) dx \right] dx.$$

\(^{35}\)The same remark hold also for the coefficients $\beta_{x^{dam}}^{x}$ and $\beta_{x^{2}}$ associated to the variables $x_{i,f}^{dam}$ and $x_{i,f}^{2}$. 

24
In these two expressions, the notation $E_{x \sim \mathcal{G}^{CV}(.|\tilde{x}, f_{inc}^i)}$ stands for the expectation with respect to $x$ (the sum of bid-signals $bs_{i, f_{inc}^{i}}$) conditional on the identity of the incumbent, $f_{inc}^i$, and $\tilde{x}$ the realization of the incumbent’s aggregate bid-signal $bs_{i, f_{inc}^{i}}$.

Under the symmetric Gaussian structure, the distribution of $bs_{i, f_{inc}^{i}}$ conditional on $bs_{i, f_{inc}^{i}} = x$ is a normal distribution with mean $\frac{\sigma^{2}_{CV}}{\sigma^{2}_{PV} + \sigma^{2}_{CV}} \cdot x$.

Using in addition the expression of $\tilde{bs}_{i, f_{inc}^{i}}(w_i)$ (given in (3)), then for $l = 1$, (13) and (14) reduces to

$$CF_{i, f_{inc}^{i}}^{PV}[1] = \frac{\sigma^{2}_{PV}}{\sqrt{\sigma^{2}_{PV} + \sigma^{2}_{CV}}} \cdot \Phi\left(\frac{\log(w_i) - A_{i, f_{inc}^{i}}}{\sqrt{\sigma^{2}_{PV} + \sigma^{2}_{CV}}}\right)$$

and

$$CF_{i, f_{inc}^{i}}^{CV}[1] = \frac{\sigma^{2}_{CV}}{\sigma^{2}_{PV}} \cdot CF_{i, f_{inc}^{i}}^{PV}[1].$$

In (15), $CF_{i, f_{inc}^{i}}^{CV}[1]$ is a non-linear function of $log(w_i)$, and the variables $x_{i, f_{inc}^{i}}^{w}$ and $z_{i, f_{inc}^{i}}^{w}$. If we limit ourselves to the sample with incumbents and where the incumbent is the winner, then the variables $x_{i, f_{inc}^{i}}^{w}, x_{i, f_{inc}^{i}}^{dum}$ and $CF_{i, f_{inc}^{i}}^{CV}[1]$ will be of full rank, which implies that the associated coefficients $\beta^2_{x}, \beta^2_{z_{inc}}$, and $\gamma^{CV}$ are identified. On the whole, case 1 and 2 are enough to get the full rank condition and thus to identify the augmented performance equation (10).

Cases 3 and 4 are much more tedious and detailed in Appendix A.2. Below we give some details to get a better understanding of the difficulties to estimate the control functions.

**Case 3:** Under the symmetric Gaussian structure and for $l = 1$, we get the following expression for the control functions:

$$CF_{i, f_{inc}^{i}}^{PV}[1] = \sigma_{PV} \cdot \frac{\log(w_i)^{-A_{i, f_{inc}^{i}}}}{\sqrt{\sigma^{2}_{PV} + \sigma^{2}_{CV}}} \cdot \Phi\left(\frac{\log(w_i)^{-A_{i, f_{inc}^{i}}}}{\sigma_{PV}}\right)$$

and

$$CF_{i, f_{inc}^{i}}^{CV}[1] = -\frac{\sigma^{2}_{CV}}{\sqrt{\sigma^{2}_{PV} + \sigma^{2}_{CV}}} \cdot \Phi\left(\frac{\log(w_i)^{-A_{i, f_{inc}^{i}}}}{\sqrt{\sigma^{2}_{PV} + \sigma^{2}_{CV}}}\right),$$

where $U_{i, f_{inc}^{i}}^{A}$ is the third term appearing in $\tilde{bs}_{i, f_{inc}^{i}}(W_i^{r})$ as defined in Section 3.3. In Appendix A.2 we show that the term $U_{i, f_{inc}^{i}}^{A}$ equals

\(36\) Note that $x_{i, f_{inc}^{i}}^{dum} = 1$ in this subsample.

\(37\) This term can be interpreted as the amount by which the auction winner shifts his cutoff bid-signal to account...
In cases 1 and 2, two effects play in favor of the full rank condition: i) control terms are non-linear in the observable covariates \( x_{i,f} \), \( z_{i,f} \) associated to the winner (i.e. for \( f = f^w_i \)), ii) control terms depend on the variables \( z_{i,f^w} \) which play thus role of instrument. However, we do not exclude that there are no such auction covariates: Those covariates are no used for identification, they should rather be viewed as a plus from an estimation perspective to reduce colinearity issues.

In case 3, we see that there is another force that reduces the colinearity between the original regressors and the control terms: the later terms depend not only on the covariates of the winner but also of the incumbent, i.e. \( x_{i,f^w} \), \( z_{i,f^w} \) (through \( A_{i,f^w} \)) which play thus the role of an instrument for selection.

A new difficulty also arises: in cases 1 and 2, the control term \( CF_{i,f^w}^{PV}[1] \) depends on the parameter \( \text{theta} \) and not only through the term \( (1 - \text{theta}) \cdot \sigma^2_{CV} \) as we can see form the expression of \( U_{i,f^w}^{A,inc} \). This is potentially an issue since \( \text{theta} \) is no identified when \( \sigma_{CV} = 0 \). However, a Taylor expansion reveals that

\[
U_{i,f^w}^{A,inc} = -\sigma^2_{CV} \cdot \frac{\phi\left(\frac{[\log(W^r_i) - A_{i,f^w}^{inc}]}{\sqrt{\sigma^2_{CV}}}\right)}{\phi\left(\frac{[\log(W^r_i)]}{\sqrt{\sigma^2_{CV}}}\right)} + \frac{(1 - \theta)}{2} \cdot \sigma^2_{CV} + O(\sigma^4_{CV}).
\]

This suggests that our methodology could work well even if \( \sigma_{CV} \) is small as confirmed by our simulations.

**Case 4:** When the final price is strictly above the reserve price, \( w_i > W^r_i \), and the winner is not the incumbent, \( f^w_i \neq f^inc_i \), the calculations of the control functions are relatively complex. The difficulty comes mainly from the fact that the bid-signals of the winner and the incumbent are not independently drawn conditional on \( \mathcal{F} \). This is so because the bidding strategy of the non-incumbents depends on the moment when the incumbent has dropped out from the auction, and this moment is not observed by the econometrician. Detailed calculations are therefore relegated to the Appendix. Nevertheless, if the econometrician could observe the moment where the incumbent has quit the auction, then we obtain the following expressions for the control functions (under the symmetric Gaussian structure and \( for l = 1 \)) denoted now \( CF_{i,f^w}^{PV}(b_i^{inc})[1] \) and \( CF_{i,f^w}^{CV}(b_i^{inc})[1] \) where \( b_i^{inc} \) denotes the drop-out bid of the incumbent and where \( b_i^{inc} = NP \) if the incumbent does not enter the auction.

for the fact that the incumbent did not enter the auction at the reserve price \( W^r_i \). Note that \( U_{i}^{A} \leq \frac{(1 - \theta)}{2} \cdot \sigma^2_{CV} = A_{i,f^inc}^{inc} - A_{i,f^inc}^{inc} \), the value of the signal shifter when the reserve price goes to infinity or equivalently if the incumbent were not eligible to bid.
of the winner and of the incumbent (see the Appendix), and where the term $\tilde{b}_{f,w}(w_i, \tilde{b}_{f,inc}(b_{i,inc}^*))$ can be shown to be linear in $\log(w_i)$ and the covariates of the winner and of the incumbent (see the Appendix), and

$$CF_{i,f}^{PV}(b_{i,inc})[1] = \sigma_{PV} \cdot \frac{\phi\left(\frac{\tilde{b}_{f,w}(w_i, \tilde{b}_{f,inc}(b_{i,inc}^*))}{\sigma_{PV}}\right)}{1 - \phi\left(\frac{\tilde{b}_{f,w}(w_i, \tilde{b}_{f,inc}(b_{i,inc}^*))}{\sigma_{PV}}\right)}$$

if $b_{i,inc}^* \neq NP$, and $CF_{i,f}^{PV}(NP)[1] = \sigma_{PV} \cdot \frac{\phi\left(\frac{\tilde{b}_{f,w}(w_i^*)}{\sigma_{PV}}\right)}{1 - \phi\left(\frac{\tilde{b}_{f,w}(w_i^*)}{\sigma_{PV}}\right)}$.

The novelty of this last case is that the control terms depart from Mills ratios. On the one hand, if we could observe the incumbent’s drop-out bid, then $CF_{i,f}^{CV}(b_{i,inc})[1]$ (for $b_{i,inc}^* \neq NP$) is a linear function of the covariates of the incumbent (instead of a non-linear one as in cases 2 and 3). This difference reflects also that in this case, we are able to recover perfectly the incumbent’s bid-signal and get a less noisy control term. On the other hand, if we could not observe the incumbent’s drop-out bid, then the correct control terms will correspond to a weighted-average of the terms in (18) and (19) and are thus no longer Mills ratios. How to compute the associated weight is developed in Appendix.

4.3 Estimation

Our estimation method follows our identification argument and involves thus two stages: first, the primitives of the auction model are estimated through maximum likelihood estimation; second, the parameters of the performance equation are estimated by OLS with the estimated control functions added as controls.

The likelihood function of the auction data is given in the Appendix. Although the distribution functions $G_i^{CV}$ and $G_{f,i}^{PV}$, $f = 1, \cdots, F$, are non-parametrically identified (Section 4.1), we adopt here a parametric approach and thus assume that these functions are known up to a parameter belonging to $R^d$ for some $d \geq 1$. These parameters are estimated in the first stage together with the scalars $\theta$ and $\beta_{f}^*$, $f = 1, \cdots, F$, and the vectors $\beta_{x}^*$ and $\beta_{z}^*$, which altogether completely characterize how firms bid as a function of their bid-signals, and this for any worker $i$, given the set of observable covariates $x_{i,f}, z_{i,f}$, $f = 1, \cdots, F$. We will assume here that the identity of the second-highest bidder is observed by the econometrician, as is the case in our dataset.

Some remarks on the likelihood function. 1) In writing down the likelihood, one should carefully pick the appropriate density or distribution functions of the cutoff bid-signals associated
to observed bidding behavior (depending on whether there is an incumbent and if there is one, whether the incumbent is the winner or the second-highest bidder, and also on whether the worker remains unsold) and which reflects the different bidding regimes A, B, C when there is an incumbent, but otherwise its structure is relatively simple and resembles some of the likelihoods derived in other papers (see e.g. Baldwin et al. (1997)).

2) Depending on the specific parametric distribution function chosen for the bid-signals, it may occur that the support of the observables depends on the vector of parameters say $\mu$ or more generally that the conditional density suffers from discontinuities with respect to $\mu$. This would violate the regularity conditions required to derive the usual.

Our second stage consists in estimating the performance equation

$$y_{i,f} = \beta_{f,w} + \beta_{x} \cdot x_{i,f} + \tau \cdot \log(w_{i}) + \gamma_{PV}^{PV} \cdot CF_{i,f}^{PV} + \gamma_{CV}^{CV} \cdot CF_{i}^{CV} + \text{error}_{i,f},$$

using the performance data. Here $CF_{i,f}^{PV}$ and $CF_{i}^{CV}$ are the estimated control functions, i.e., the expressions one gets after replacing all unknown parameters entering the control functions by their first-stage estimates. The parameters $\beta_{f}, \beta_{x}, \tau, \gamma_{PV}^{PV}$, and $\gamma_{CV}^{CV}$ are estimated by OLS.

To account for the fact that in (20) we substituted the estimated control functions for their unknown counterparts, the standard errors are obtained by a percentile bootstrap method based on 1,000 bootstrapped samples.

5 Related literature and identification variants

5.1 Related literature on auctions and some comments on identification more generally

The vast majority of the applications of the structural econometrics of auctions involves pure private value models with risk-neutral bidders. Our auction model and its identification from bidding data only are novel due to two main ingredients: 1) Bidders’ payoff functions are no longer quasi-linear in the auction price (if $\theta \neq 0$); 2) The model involves interdependent values with asymmetric bidders. Such ingredients have already being addressed to some extent in the literature as discussed next. Note also that our model involves multi-dimensional signals. This is not innocuous from an econometric perspective. However, from a bidding perspective, our auction model is equivalent to one with single-dimensional signals: as in Goeree and Offerman (2003)’s interdependent values model with bi-dimensional private signals, the bidding incentives of bidder $f$ for worker $i$ depend on his private information $s_{i,f}$ only through his single-dimensional bid-signal $bs_{i,f}$.

In the existing literature, departures from the quasi-linear paradigm arise typically from risk aversion which is known to cause important identification problems (Guerre, Perrigne, and...
Vuong (2009)), but also possibly from contingent payment auctions (like equity auctions) where monetary transfers depend on some ex-post realization (as in Battacharyya, Ordin, and Roberts (2018)). In our setup, the departure from the quasi-linear paradigm is captured by the parameter $\theta$. Once the distributions of the bid-signals of the private and common value components of a given incumbent are identified, we have shown that $\theta$ is identified by comparing the equilibrium bid distributions of a given non-incumbent in the samples with and without the given incumbent, i.e. in a sample where $\theta$ drives bidding behavior with one where its plays no role. More precisely, we argued in Section 4 that any quantile of the bid distribution of a non-incumbent in the sample with an incumbent allows to identify $\theta$: this parameter is thus largely over-identified. This suggests that much more general forms could be used for the incentive effect (instead of the linear specification $\tau \cdot \log(w)$). Extending our model in this direction, and studying the possible identification issues that the may arise, is much beyond the scope of this paper and left for further research.

It is well-known that interdependent values raise important issues. On the one hand, equilibrium existence can fail in the second-price auction or in the English auction as illustrated by Jackson (2009), contrary to the first-price auction where Reny and Zamir (2004) obtain a very general positive result. On the other hand, Laffont and Vuong (1996) argue that without exclusion restrictions, any bidding data generated by an interdependent value model can be rationalized by a pure private value model.

Most auction models with interdependent values impose that bidders are fully symmetric (see e.g. Milgrom and Weber (1982)'s affiliated value model which has been applied to oil and gas lease auctions by Hendricks, Pinkse, and Porter (2003)). Goeree and Offerman (2003) develop a symmetric model where each bidder receives a bi-dimensional private signal who reflects respectively a pure private and a pure common value component. By contrast, in our model, only the incumbent firm receives a signal about the common value of the worker, an ingredient we took from Engelbrecht-Wiggans et al. (1983)'s seminal asymmetric pure common value model (see also Hendricks et al. (1994) for a generalization with a random reserve price). On the whole, our model can then be viewed as an hybridization of Engelbrecht-Wiggans et al. (1983) with Goeree and Offerman (2003) while allowing for general forms of asymmetries between the distributions of the private value components across bidders.

Extending our model to allow for the information about the common value component to be dispersed (possibly asymmetrically) among multiple bidders would be of primary interest. In a related vein, Lu and Perrigne (2008) identify jointly bidders’ valuation distribution and their risk aversion -both in a non-parametric way- thanks to the observation of bidding data for both the English auction (where equilibrium strategies do not depend on bidders’ risk aversion) and the first price auction (where more risk-averse bidders bid more aggressively). In a two-stage sequential auction setup where the first auction is a first-price auction and the second an English auction, Kong (2018) deals with general multi-object preferences and risk aversion but stick to pure private values with independent signals across bidders. Pinkston (2009) who develops such a model for a labor market framework with an incumbent, maintains the assumption that the value of a worker is common across all firms. Said differently, his model does not contain any
However, it would raise important challenges both in terms of equilibrium existence and identification, at least without further restrictions on the signals distributions. Those issues could be circumvented by imposing some parametric restrictions.\textsuperscript{40}

To the best of our knowledge, we are the first to deal with the non-parametric identification of a model with interdependent values and asymmetric bidders. Laffont and Vuong (1996)’s impossibility result can be circumvented with richer data, e.g. when the value of the good for sale is observed perfectly ex-post (Hendricks, Pinkse, and Porter (2003)).\textsuperscript{41} We identify and estimate our auction model based on bidding data only and an exclusion restriction on the signal distributions for the samples with and without incumbents. While we do not use the post-auction data for this first stage, it may be profitable to use it to refine our estimators.

5.2 Identification of our auction model more broadly

The aim of this subsection is to sketch how identification could be reached either with alternative exclusion restrictions, with different commonly observed bidding rules, or last in the English auction if we benefit from additional information on the bidding dynamics. The list is far from being exhaustive. Our main message is that we deliberately place ourselves in the most difficult case (at the cost of relying on arguably strong assumptions).

**Exogenous variation of covariates** The identification argument we present does not rely on exogenous variations of the covariates exactly as in the bulk of the literature on the econometrics of auctions. Indeed, for any set of covariates, we can identify the primitives of interest. Nevertheless, the exclusion restriction A8 could be substituted by an assumption where bidder-specific covariates vary across auctions. As explained in Athey and Haile (2008), such variations allow identification in a much simpler manner if we are prepared to assume that those variations are exogenous. In particular, samples without an incumbent can be substituted by samples with an incumbent having poor covariates ($A_{i, f^{inc}}$ going to minus infinity) so that the probability that the incumbent enters the auction goes to zero.

**Reserve price** If the auction rules involve a binding reserve price, then Roberts (2013) shows that variations in this reserve price allow to deal with unobserved heterogeneity in the English auction when there is a monotonic relation between the reserve and the “quality” of the good for sale that is observed by bidders but not the econometrician. Note that exogenous variations in the reserve price could help identification and circumvent Laffont and Vuong (1996)’s impossibility result. Nevertheless, reserve prices may also raise some issues if we wish to identify private value component.

\textsuperscript{40}See Hong and Shum (2003) and Heumann (2017) for interdependent values models involving a Gaussian information structure and with respectively single-dimensional and multi-dimensional signals. Relatedly, Weiergraebener and Wolf (2018) develop an empirical analysis of a generalization of Goeree and Offerman (2003)’s model but without any formal guarantees neither in terms of equilibrium existence nor in terms of identification.

\textsuperscript{41}Bhattacharya, Ordin, and Roberts (2018) rely also crucially on the observation of post-auction data about an ex-post investment made by the winner and which indirectly reveals critical information about the ex-post realization.
the signal distribution over its full support. This potential issue can be solved by sufficient variations in the covariates.\textsuperscript{42} Furthermore, rather than the full identification, what matters for the second step is to be able to control for the endogeneity biases. If we are in a pure private value setup, then we do not care about identifying the distribution \( G_{f}^{PV}(\cdot) \) for signals \( \bar{b}_{f} \) that are below \( \bar{b}(W_{i}^{*}) \) because we need to identify \( E[bs_{i,fw}^{PV}|\vartheta] \) only up to a constant.

**Other auction formats** In other environments, especially in procurements, the auction format is rather a sealed bid first price auction rather than the English auction. If we observe only the winning bid, we can easily adapt our methodology. Indeed, we should also benefit from a much more precise estimation of the control term \( E[bs_{i,fw}^{PV} | \vartheta] \) insofar as instead of learning that the winner’s bid-signal is above some cutoff we can identify his bid-signal directly from its bid. If we observe multiple bids, then this information would be helpful from an identification perspective, either to relax slightly our exclusion restrictions, or allow for unobserved worker heterogeneity as in Li, Perrigne, and Vuong (2000) or Krasnokutskaya (2011).

**Additional information on the auction dynamics** For identification, we have considered the most difficult case in the English button auction by not assuming that the identity of the second highest bidder is observed, neither that other drop-outs are observed. In our data set, we do actually observe all drop-outs.

### 6 Simulation study

The aim of the simulation study is to show how our methodology performs using small sample sizes (we pick \( N = 300 \) or \( N = 1,000 \) auctions). The parameters of our Monte Carlo protocol have been chosen as follows. As in our empirical application, we set the total number of bidders to \( F = 8 \). Throughout, it is assumed that the set of potential participants coincides with the full set of eight bidders. In each simulated data set, half of the sample is composed of auctions without incumbents while the other half is made up of auctions with an incumbent (picked randomly). In all simulations we fix \( \lambda = 1 \), while \( \tau \) can take the values \(-0.8, -0.5, 0, 0.5 \) and 0.8 (these are hence also the values taken by \( \theta \)). As in the illustrative example of Section 4.2, we consider symmetric Gaussian structures. In addition, we assume that \( bs_{i,fw}^{id} = bs_{i}^{co} = 0 \) for all \( i \) and \( f \), implying that \( d_{CV}^{CV} = d_{f}^{PV} = 1 \), which in turn means that the coefficients associated with the two control functions in (10) equal \( \gamma_{CV}^{CV} = \gamma_{f}^{PV} = 1 - \tau \) for all \( f \). For all simulations we set \( \sigma_{PV} = 1 \), while \( \sigma_{CV} \) take the values 0, 1 and 2 (\( \sigma_{CV} = 0 \) corresponds to the private value case).

The vectors of covariates \( x_{i,f} \) and \( z_{i,f} \) are both single dimensional. The worker/firm character-
istic \( x_{i,f} \) is assumed to be the same across all bidders, but the auction variable \( z_{i,f} \) does vary with \( f \). Both \( x_{i,f} \) and \( z_{i,f} \) are assumed to be i.i.d. (across \( i \) and \( f \)) according to a centered normal distribution with variance equal to 1. We set \( \beta_f = 0 \) for all bidders and \( \beta_x = 1 \), while the noise \( e_{i,f} \) is assumed to be distributed according to a centered normal distribution with variance equal to 1.\(^{43}\) We also set \( A_{i,f} = x_{i,f} + z_{i,f} \), which amounts to choosing the following parameter values appearing in \( \bar{V}_f(x_{i,f}, z_{i,f}) \): \( \alpha_f = -\frac{1}{2}, \alpha_x = -\tau \) and \( \alpha_z = 1 - \tau \). Finally, the reserve price \( W_{i}^r \) is set to \( 2 \cdot \frac{\lambda}{1-\tau} = \frac{2}{1-\tau} \). This guarantees that the probability of entering the auction remains constant for the incumbent when \( \tau \) varies.

Table 1 reports the mean estimates (over 1,000 Monte Carlo replications) of the parameters \( \tau \) and \( \beta_x \) for three methodologies: 1) standard OLS (estimation of the performance equation without the two control terms); 2) A variant of our methodology (referred to as the PV methodology), which consists in estimating a simplified version of our model and the corresponding parameters as if the incumbents do not have superior information relatively to non-incumbents;\(^{44}\) 3) Our general methodology described in Section 4 and referred to as the CV methodology.

Table 2 reports the estimated lower and upper bounds of the 95% confidence intervals for \( \tau \) and \( \beta_x \). We estimate these bounds using the Warp-Speed method developed by Giacomini, Politis, and White (2013). This method consists in drawing, from each Monte Carlo replication sample, a single bootstrap sample of size \( N \), and then in calculating the second-stage estimates of the two parameters of interest based on both samples. For each parameter, we produce then 1,000 CIs which are obtained by adding to each Monte Carlo estimate the 2.5% and 97.5% percentiles of the distribution of the difference between the Monte Carlo estimate and its associated bootstrap estimate. Our estimate of the lower (resp. upper) bound is then simply the average of the lower (resp. upper) bounds of these CIs. The major advantage of the Warp-Speed method is that only a single bootstrap sample is required for each replication, thereby drastically reducing the computational cost of Monte Carlo experiments involving bootstrap estimators.\(^{45}\) In the Appendix the Warp-Speed technique is briefly summarized. Also reported in Table 2 are estimates of statistical power. We calculate the power of the test of the null hypothesis \( H_0 : \tau = 0 \) (resp. \( H_0 : \beta_x = 0 \)), against the bilateral alternative, as the fraction of times zero does not belong to the 1,000 CIs for \( \tau \) (resp. \( \beta_x \)), given that the data are generated under a particular value of \( \tau \) (resp. \( \beta_x \)). The results in this table are only given for the CV

---

\(^{43}\)These choices ensure that one half of the sample variation in the simulated performance measure is explained by \( x_{i,f} \), about the same fraction as in our real data.

\(^{44}\)The likelihood maximized at the first stage is thus the one corresponding to a standard PV model with all bidders having cutoff bid-signals of the form (6). Compared to our model, it is as if \( \sigma_{CV} = 0 \) while the parameter \( \theta \) is no longer present in the likelihood. The performance equation estimated at the second stage only includes the control function \( CF_{PV}^i \) (the control \( CF_{CV}^i \) is imposed to equal 0), defined in (12) with \( \sigma_{CV} = 0 \) under the Gaussian specification.

\(^{45}\)For instance, the power and confidence interval computations for each parameter set in our study (each column of Table 2) takes approximately one day with the Warp-speed method (this corresponds to one bootstrap draw). Standard bootstrap methods, requiring anywhere between 500-1000 draws, would have exponentially increased the computational time from one day for each parameter set to possibly 500-1000 days.
methodology. We see from Table 1 that, as predicted in Section 2, the OLS estimator of $\tau$ is upwards biased while the one of $\beta_x$ is biased toward zero. The bias is very substantial for both parameters, especially when $\tau$ takes small values.\footnote{This is as expected from our specification where the parameters $\gamma_{PV}$ and $\gamma_{CV}$ both decrease in $\tau$ and vanish when $\tau = 1$, so endogeneity matters more when $\tau$ is small and becomes negligible as $\tau = \theta$ goes to 1.} For instance, when $\tau = -0.8$ and $N = 1,000$, the mean OLS estimate of $\tau$ ranges between -0.23 ($\sigma_{CV} = 0$) and 0.17 ($\sigma_{CV} = 1$), while the mean OLS estimate of $\beta_x$ (recall that its true value is 1) varies between 0.09 ($\sigma_{CV} = 1$) and 0.43 ($\sigma_{CV} = 0$). We also see that the smallest biases are always obtained when $\sigma_{CV} = 0$. For $\tau \geq 0$, the biases increase in $\sigma_{CV}$, while for the case $\tau = -0.5$ and $-0.8$ the strongest biases are obtained for the intermediary value $\sigma_{CV} = 1$.

The PV estimator performs well when $\sigma_{CV} = 0$. This does not come as a surprise because precisely for this value the data are generated under the PV paradigm, and the PV methodology then provides consistent estimators. We see from Table 1 that the results are not satisfactory when $\sigma_{CV} > 0$. For instance, if $\tau = -0.5$ and $\sigma_{CV} = 1$, the mean estimate of $\tau$ (resp. $\beta_x$) is around -0.21 (resp. 0.74) for both sample sizes, a relative bias of approximately 58% (resp. 26%). Nevertheless, the PV methodology produces much better results than OLS. Looking for example at the estimates of $\tau$, the PV method reduces roughly half (resp. two thirds) of the bias of the OLS estimator when $\sigma_{CV} = 2$ (resp. $\sigma_{CV} = 1$).

Consider finally the results for the CV estimator. Its performance is comparable to the performance of the PV estimator when $\sigma_{CV} = 0$. This is quite surprising since there is no guarantee that our CV methodology is consistent under the PV paradigm (contrary to the PV methodology). At the same time, Table 1 shows that the CV estimator clearly outperforms the PV estimator in the case $\sigma_{CV} > 0$. Indeed, when $\sigma_{CV} = 1$, the CV estimator of $\tau$ is perfectly consistent for any value of the wage effect; the parameter $\beta_x$ is also well estimated although there is a small upwards bias for negative values of $\tau$. When $\sigma_{CV} = 2$, the CV estimator of $\tau$ (resp. $\beta_x$) is slightly downwards (resp. upwards) biased, but the the biases are much smaller than those produced by the PV estimator.

Table 2 shows that our tests of $H_0 : \tau = 0$ and $H_0 : \beta_x = 0$ have high power: the probability of correctly rejecting the null hypothesis $H_0 : \tau = 0$ fluctuates between 69 and 100% when $N = 300$, and equals at least 98% when $N = 1,000$; the probability of correctly rejecting $H_0 : \beta_x = 0$ is at least equal to 98% in the smaller samples, and always equal 100% in the larger ones. Note that the power results reported in the upper panel and under the columns headed $\tau = 0$ actually give the size of the test of $H_0 : \tau = 0$ against the two-sided alternative. As indicated in the table, the size is adequate when $\sigma_{CV}$ equals 0 or 1 (it then varies between 0.04 and 0.06), but seems a bit too large when $\sigma_{CV} = 2$ (0.10 if $N = 300$ and 0.24 if $N = 1,000$). This last result can be explained by the fact that, as mentioned above, the CV estimator of the wage effect is slightly biased for this relatively high value of the standard deviation.
Finally, we see from Table 2 that the estimated lower-bounds and upper-bounds of the CIs are generally symmetrically distributed around the true values (especially those corresponding to $\beta_x$). This symmetry is weaker when $\sigma_{CV} = 2$ because the CV estimator is then slightly biased. In this case the bounds are symmetrically distributed around the estimated mean values. We observe that, as expected, the CIs are tighter for the larger samples than for the smaller ones. We also observe that the CIs for $\beta_x$ become tighter as $\sigma_{CV}$ gets smaller. Contrastingly, there is not a clear-cut relationship between the CIs for $\tau$ and the standard deviation $\sigma_{CV}$: for $\tau = -0.8, -0.5, 0$, the CI is tighter as $\sigma_{CV}$ gets smaller, while the reverse holds for $\tau = 0.5, 0.8$. Intuitively, there are two countervailing forces when $\sigma_{CV}$ gets larger: on the one hand, the selection effect becomes more important and introduces additional noise in the performance equation (formally, the variance of the residual $\xi_{i,fw}$ gets larger), reducing the precision of the estimates. On the other hand, $\bar{CF}_{i,fw}$ becomes a less noisy estimator of the control term $CF_{i,fw}$ ($\theta$ is no longer identified when $\sigma_{CV} = 0$ and thus $CF_{i,fw}$ will be poorly estimated when $\sigma_{CV}$ is small), thereby instead augmenting precision. Depending on the parameter being estimated (and the true value of $\theta$), one of these forces dominates the other.

7 Data and empirical application

7.1 Tournament and player performance

The Indian Premier League (IPL) is an annual cricket tournament where teams compete by playing matches in a double round-robin format. At the end of this first stage, the four best ranked teams compete in a playoff to determine the final winner of the tournament. In our empirical analysis we focus on the 2014 IPL because it represents a year in which major player auctions were held before the tournament, whereby players were (re)allocated to teams. In that year, eight teams competed in the tournament and each team played between 14 and 16 matches depending on whether it qualified in the playoff.

A cricket match is played over a fixed time period (three hours in the IPL) between two teams consisting of 11 players who are selected from the team squads. Cricket players are categorized into four categories: batsman, bowler, wicket keeper and all-rounder. The unique feature of cricket is that, unlike most other team sports, a large component of overall team performance depends on individual specific performances. Since player skills are highly specialized, it is possible to observe a large set of individual measures of performance that are idiosyncratic and largely independent of how other team members perform. On average, we observe 115.6 performance measures per team in 2014 (note that some players in the selected squad may not perform).

From the individual player performances we construct a composite performance measure. It
is derived from various, batting and bowling statistics observed for each player during the tournament. We award points for each statistic accumulated by each player across every game. The way that a player accumulates points and the construction of our composite performance measure is described in detail in Appendix A.3. We also give there additional information on the rules of the cricket game.

**Data sources:** We obtained performance data on all matches played in the tournament from www.espncricinfo.com. All data on player auctions, described in the following section, were manually compiled from the recordings and minutes of the (publicly-broadcast) auction proceedings.

### 7.2 Player auctions

Beginning in 2008, once every three years, the IPL organizes auctions to (re)allocate players to teams. This centralized market is the main opportunity for teams to hire their players. The format of sale consists of a sequential procedure whereby players are sold sequentially, one after the other, through a series of English (or ascending) auctions with public reserve prices.\(^\text{48}\) In each of those independent auctions, the winning bid represents the player's salary for the IPL tournament. Furthermore, the auctions were the only way to hire new players, implying in particular that those remaining unsold in the auctions do not participate in the tournament.

Prior to the sale, each player is assigned a reserve price that represents the price at which bidding for a player starts. The reserve price is broadly determined by the auctioneer based on a variety of factors, primary among them being the player's past performance. The auctioneer also arranges the players into different ‘sets’ by their cricketing speciality, popularity, and, to some extent, their reserve price. The sale of players proceeds according to a predetermined sequence of these sets. The composition of the sets and the sequence in which they are placed in the auction are announced ex ante. Within each set, the order in which players are auctioned is determined by random draws in the format of a lottery.

The teams faced a set of explicit rules with regard to both team composition and bidding behavior. These rules play an important role in determining some constraints that bidders face whilst bidding. In 2014 there were three types of rules. 1) **Spending cap:** in order to encourage a balanced competition, the organizers imposed a spending cap on the total amount that any bidder was allowed to spend in the auctions. The spending cap allocated to a bidder depends on the number of players retained by the team from its previous year’s squad (the less players retained, the higher the cap). In 2014, teams were allowed to retain a maximum of five players from their previous year's squad, and the spending cap varied from 245 to 700 Millions of Rupees.

---

\(^\text{47}\)The minutes of the live auction proceedings were obtained from ESPN-Cricinfo (http://www.espncricinfo.com/indian-premier-league-2014/content/story/718095.html).

\(^\text{48}\)In 2014, Seven different reserve prices were used from 1 up to 20 Millions of Rupees.
2) **Overseas player quota:** to ensure a sufficient number of native players in the tournament, the organizers imposed a maximum limit of 9 on the number of overseas (non-Indian) players in any team. 3) **Right-to-Match (RTM) option:** depending on the number of retained players, each team received from the organizers between 1 to 3 so-called Right-to-Match cards. A RTM card allowed a team to buy back a player from its previous year’s squad by matching the player’s winning bid when he was sold at the auction.\(^{49}\)

Each player was auctioned through an English auction where teams were invited to challenge the temporary winner by raising their paddle to indicate their willingness to buy the player at the current price plus a predetermined increment. However, our analysis abstracts from the bid increments and proceeds as if the bidding data is generated from an English button auction (see Section 3.1) as in almost all the literature.\(^{50}\) If the player was not RTM-eligible, then the provisional winner was declared the final winner and the player was sold to this team at a price equal to its bid. If the player was RTM-eligible, his team from the previous year had the option to use one of its RTM cards and match the winning offer to buy-back their player.\(^{51}\)

### 7.3 Descriptive statistics

In 2014, a total of 317 players were auctioned and 8 teams participated. Out of these 317 players, 105 received bids at or above the reserve prices and were actually sold. For all players (including unsold players) we know a number of characteristics: their nationality, their cricket speciality, and whether they are a so-called newcomer.\(^{52}\) We record, for every auction, whether in the previous year the player was playing for one of the 8 teams, and, if this is the case, the identity of the player’s previous team. Using the terminology previously used in the paper, this bidder is called the incumbent and the corresponding team is referred to as the incumbent team. We observe how players are pooled into sets, the auction sequence between sets and within sets. We also observe in the data all reserve prices attached to the players, and, for those sold, the composite performance measure defined in the Appendix. For each auction we observe all submitted bids (i.e., all prices at which teams raised their paddles) together with the identities of the corresponding bidders, and the identity of the team who has used a RTM card (if any).

Tables 3 and Table 4 present summary statistics from the perspective of teams and players, respectively. The upper panel of Table 3 shows that approximately half of all auctions (156

---

\(^{49}\)A RTM card is equivalent to what is called the “right-of-first-refusal” option in the auction literature (see Bikhchandani, Lippman, and Ryan (2005) for an analysis of second-price/English auctions with this option).

\(^{50}\)A notable exception is Haile and Tamer (2003) in a pure private value environment, but at the cost of dealing with partial identification. Lamy, Patnam, and Visser (2016) do consider increments but at the cost of being not fully structural at the bidding stage, and this for a pure private value model.

\(^{51}\)Note that a team with the possibility to exert the RTM option was allowed to bid in the auction as any of the other bidders.

\(^{52}\)A newcomer is a player who has already been called by his national team. Such players are of course already experienced, but we nonetheless use the terminology “Newcomer” as this is the official designation for such players.
out of 317) involved an incumbent. In 75 of these auctions the incumbent was eligible to use a RTM card and did use it in 17% of these cases, and in the remaining 81 auctions he did not because he did no longer possess a RTM card. As indicated in the middle panel, a team purchased on average 15 players through the auctions considered, comprising approximately of 3-4 batsmen, 6-7 bowlers, 1-2 wicket keepers, and 3-4 all-rounders. Furthermore, about 10 of the newly purchased players were Indian, and 7 newcomers. Teams bought 1.62 players trough the RTM option. The lower panel shows that teams retained on average 3 players form their previous year's squad (these players do not appear in the auctions), and were allocated a budget of nearly 5 million USD for purchasing players. On average, bidders consume 90% of their allocated budget.

Table 4 contains summary statistics on the auction data, for the full sample in the upper panel, and for the sold-players sample in the lower one. The lower panel also reports statistics on our composite measure of performance and the wage earned by the cricket players. Conditional on attracting some bidders, the average number of participants (i.e. the number of bidders having raised their paddle at some moments) is 2.4. The reserve price for sold players is found to be not significantly different compared to the sample average, in both cases it was set around 0.1 million USD. The fraction of newcomes and the within-set order of player appearance in the auctions is also similar across the two samples. The fraction of Indian players and the fraction of players eligible for RTM are, however, slightly higher in the sold-players sample. The average winning price is 0.37 million USD. There is actually a huge heterogeneity in the wages: the ratio between the highest and the lowest wage obtained in the auctions is as large as 140. Finally, we see that the performance score on average equals 23.9, and there is much dispersion in this variable as well since its standard deviation equals 14.55.

7.4 Empirical issues

In this section we introduce the empirical specifications chosen in our application. We also briefly outline how our methodology is (slightly) adapted to take into account several features of the data.

Again, we adopt a symmetric Gaussian structure. The variables we include in the vector of team/player characteristics $x_{i,f}$ are cricket-speciality dummies indicating whether $i$ is of a certain speciality, a dummy indicating whether $i$ is of Indian nationality, and a dummy indicating whether he is a newcomer. We also include the three indicator variables Incumbent present & RTM card (equal to 1 if one of the eight firms is the incumbent and a RTM card is available, 0 otherwise), Incumbent present & no RTM card (equal to 1 if one of the eight firms is the incumbent and a RTM card is not available, 0 otherwise), and Bidder is incumbent (it corresponds

53 Included are dummies for batsman and bowler. There are too few wicket-keepers in the sold-player sample to add yet another speciality dummy in $x$. 

37
to the variable $x_{i,f}^{dum}$ defined earlier). The vector of auction variables $z_{i,f}$ contains the order of sale of $i$ within the set, the remaining budget of team $f$ just before $i$ is being auctioned, and five backlog variables: # Batsman bought, # Bowlers bought, # Wicket-keepers bought, # All-rounders bought, and # Overseas players bought. Each of these variables is defined as the interaction between a variable counting the number of players of a given type already bought by $f$ prior to the auction of $i$, and a dummy indicating whether $i$ is of this type.

To facilitate the maximisation of the likelihood function, we exploit that the drop-out price of the incumbent is always observed in our data: it corresponds to the last price at which the incumbent has raised its paddle. Once we condition on this price we no longer need to calculate the integral appearing in the (analogue of the) last term of the likelihood function given in the Appendix.

The presence of the RTM option for some of the IPL auctions requires to adapt slightly our analysis. If we abstract from the fact that there is a limited number of RTM cards so that the use of a card is costly, then the optimal bidding strategy of the incumbent consists in remaining silent in the auction and then in using the RTM card if and only her bid-signal $bs_{f,inc}$ is larger than $\tilde{bs}_{f,inc}(w)$ when $w$ is the termination price. For a given non-incumbent $f$ that should infer from hiring the worker at wage $w$ that the incumbent has not used her RTM card at the final price $w$, then the cutoff bid-signal, denoted by $\tilde{bs}_f^{RTM}(w)$, that makes $f$ indifferent between hiring or not the worker given that the incumbent has not used the RTM card can be expressed by:

$$\tilde{bs}_f^{RTM}(w) = \log(w) - A_f - \frac{1}{1-\theta} \cdot \log \int_{-\infty}^{\log(w) - A_{f,inc}} e^{(1-\theta)U_{f,inc}(x)} dG_{f,inc}(x) \cdot \frac{G_{f,inc}(x)}{G_{f,inc}(\log(w) - A_{f,inc})}.$$ (21)

Consequently, accounting for RTM requires small modifications in the likelihood function and the control functions. See the online appendix for a brief sketch of how the analysis should be adapted. From the perspective of the identification of the bidding model, the RTM option would simplify a lot the argument. Concerning the bid-signal distribution of the incumbent, it is immediately identified from the decision to use or not the RTM, given that the amount at which the RTM can be used varies exogenously from the auction dynamic between non-incumbents. Furthermore, since the incumbent is inactive in the auction dynamic, then the decision of the non-incumbents whether to remain active or not depend solely on their own bid-signal. Finally given our independence assumption A1 ii), the drop-out times of the non-incumbents are independently distributed and we can apply Athey and Haile (2002) to recover the distribution of the bids of the non-incumbents.

---

54Under RTM, the characterization of a separating equilibrium as in Proposition 3.1 requires to strengthen a bit A3 to guarantee that the cutoff bid-signal is increasing in $w$. Intuitively, the strength of the informational externality should not be too large.
7.5 Estimation results

Table 5 contains our empirical results. The first column gives OLS estimates of the parameters appearing in the uncorrected performance equation (1). The numbers in brackets are the 95% CIs based on the usual OLS standard errors. The results indicate that the logarithm of the player's wage has a positive and statistically significant (at the 1% level) effect on his composite performance measure. A 10% increase in wages is associated with a performance increase of 0.6 points. Regarding the team/player characteristics (the variables included in $x$), we see that the indicators for Indian players, newcomers, and batsmen are not significant. The dummy indicating whether a player is a bowler is, however, significant and has a positive effect on performance. The three indicator variables 'Incumbent present & RTM card', 'Incumbent present & no RTM card', and 'Bidder is incumbent', are not significant. As explained in Section 2, all these estimates are potentially biased because of sample selection and omitted variables. To correct for the bias, we now apply our correction method and report the results of the two estimation stages in columns 2 and 3.

Column 2 contains the first stage results, i.e., the ML estimates of the auction primitives $\beta^*_x$, $\beta^*_z$, $\theta$, $\sigma_{PV}$, and $\sigma_{CV}$, and the 95% CIs based on the asymptotic ML standard errors. All team/player characteristics are statistically significant except the newcomer indicator: teams reduce their valuation for Indian players and increase it for certain player specialties (batsmen, bowlers).

Let us next look at the results concerning our auction variables $z$. The order of sale within the set does not significantly affect the player valuations.\(^{55}\) As expected, the coefficient associated with the logarithm of the remaining budget is positive, implying that teams bid more aggressively when they have more money to spend: a 1% increase in a team's remaining budget at any point in the auction, increases a player's valuation by 0.55%.

The five last auction variables are our backlog variables capturing past purchase behavior of bidders. Three of the backlog variables have, as one might have anticipated, negative and statistically significant impacts: a team which has already acquired a batsman (resp. bowler) reduces its valuation for an additional such player by 36% (resp. 48%); the reduction for an additional overseas player is 63%, reflecting the constraint imposed by the auction organizers on non-Indian cricket players. The variable \# Wicket-keepers bought is not significant, and \# All-rounders bought is significantly positive. All-rounders being specialized in both bowling and batting, this is the type of player most wanted and used in teams, and apparently bidders

\(^{55}\) Several papers in the empirical literature on sequential auctions have shown that the final price depends negatively on the order of sale of the goods, a phenomenon referred to as the “declining price anomaly” (see for example van den Berg, van Ours, and Pradhan (2001) and Ashenfelter and Genesove (1992)). It should be noted that, unlike these papers, we study the effect of order (and other variables) on bidders’ valuations, and not on the final auction price. Furthermore, the non-significance of order does not preclude that final prices may be declining in our data. Indeed, since the remaining budget (resp. the backlog) declines (resp. increase) over the course of the auctions, and given the estimated effects of these variables, bidders’ valuations are expected to decline within a set, which in turn should introduce a declining price trend in prices.
who have already acquired all-rounders increase their valuation for an additional one by 41%.

Finally, the lower panel of the table shows that the first stage estimate of $\theta$ is 0.77 (significant at the 1% level).

Furthermore, the estimates of $\sigma_{PV}$ and $\sigma_{CV}$ are 1.48 and 2.22, indicating that a standard deviation increase in the aggregate bid-signal of a non-incumbent (resp. the incumbent) increase its equilibrium bid by 340% (resp. 1350%).

There is a substantial degree of dispersion in the sum of bid-signals $b_{sid} + b_{sPV}$ and $b_{sc0} + b_{sCV}$.

Column 3 reports the second stage results, i.e., the estimates of all parameters appearing in the augmented performance model (20), and the coefficients $\lambda$, $d_{PV}$, and $d_{CV}$, together with CIs based on standards errors obtained by a percentile bootstrap method using 1,000 bootstrapped samples. Using our control function approach, we find that the effect of wages is still significant (albeit now only at the 10% level), but much smaller in magnitude: a 10% increase in wages leads to an increase of 0.3 performance points, that is to say only half of the effect estimated by uncorrected OLS. This confirms that, as predicted in Section 2, and in accordance with our Monte Carlo results, naive OLS estimation leads to an upwards bias of the wage effect. The fact that wages still matter in explaining performance, even after controlling for sample selection and omitted variables, is (weak) evidence in support of theories emphasizing reciprocity and fairness considerations in the labor market (see Akerlof (1982) and Akerlof and Yellen (1990)).

The estimated effects of the team/player characteristics tend to be of the same sign as the OLS estimates reported in column 1. The implications of the significance tests do not change much either compared to those reported earlier. A notable exception is the variable indicating whether the bidder is the incumbent: its estimated effect has sharply declined relatively to the naive OLS estimate, and the variable is now statistically significant (at the 5% level). In line with what we predicted in 2, OLS thus indeed leads to an estimated impact of this variable which is biased towards zero. Note that, comparing columns 2 and 3, the incumbent indicator also happens to be the only variable for which the estimated coefficients of $\beta_{x}^{*}$ and $\beta_{x}$ are of different signs (newcomer is the other exception but this variable is not significant in both cases): while players perform less well when they are re-hired by their previous employer, the incumbent nonetheless values such players more highly. For incumbents such players remain popular for other reasons than their contribution to performance.

From the lower panel of the table we see that the coefficient on the control function $CF_{CV}$ is statistically significant (at the 1% level), but the one associated with $CF_{PV}$ is not. The fact that one of our control functions is significant confirms that endogeneity is an issue in our application. Finally, our estimate of $\lambda$ is 0.20 and we can reject the null hypothesis that this parameter is zero, but only at the 10% level. The estimate of $d_{PV}$ (resp. $d_{CV}$) is -0.73 (resp. 5.50), and we cannot (resp. weakly) reject the corresponding null hypothesis.
8 Conclusion

Points to address in conclusion:

- However, when the matching is not (or can not be reasonably modeled as) the outcome of an auction process or if the equilibrium (if any) of such an auction game can not be well characterized, such structural matching models could probably be substituted to our first stage auction model in order to build some control functions. The main drawback is that the stability assumptions used in this literature are a bit like a black box.
References


Table 1: Monte Carlo results: Mean estimates of $\tau$ and $\beta_x$

| $\tau$ | $\sigma_{CV}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| $\\hat{\tau}$ | $N=300$ | | | | | | | | | | | | | | | | | | | | | | | |
| OLS | -0.22 | 0.16 | -0.09 | -0.02 | 0.33 | 0.20 | 0.31 | 0.59 | 0.63 | 0.64 | 0.82 | 0.88 | 0.85 | 0.95 | 0.97 |
| PV | -0.78 | -0.45 | -0.41 | -0.48 | -0.21 | -0.09 | 0.01 | 0.19 | 0.36 | 0.49 | 0.61 | 0.73 | 0.78 | 0.85 | 0.91 |
| CV | -0.82 | -0.81 | -0.92 | -0.53 | -0.50 | -0.62 | -0.02 | 0.01 | -0.11 | 0.48 | 0.49 | 0.39 | 0.79 | 0.78 | 0.71 |

| $\beta_x$ | $N=300$ | | | | | | | | | | | | | | | | | | | | | | | |
| OLS | 0.42 | 0.10 | 0.34 | 0.52 | 0.23 | 0.35 | 0.68 | 0.46 | 0.43 | 0.84 | 0.72 | 0.66 | 0.92 | 0.88 | 0.83 |
| PV | 0.96 | 0.68 | 0.63 | 0.97 | 0.74 | 0.61 | 0.99 | 0.85 | 0.68 | 1.00 | 0.93 | 0.82 | 1.01 | 0.98 | 0.93 |
| CV | 1.05 | 1.08 | 1.14 | 1.04 | 1.06 | 1.12 | 1.03 | 1.02 | 1.09 | 1.01 | 1.01 | 1.06 | 0.99 | 1.01 | 1.02 |

| $\sigma_{CV}$ | $N=1,000$ | | | | | | | | | | | | | | | | | | | | | | | |
| OLS | 0.43 | 0.09 | 0.35 | 0.52 | 0.22 | 0.35 | 0.68 | 0.46 | 0.44 | 0.84 | 0.73 | 0.66 | 0.93 | 0.89 | 0.84 |
| PV | 0.96 | 0.70 | 0.61 | 0.97 | 0.74 | 0.61 | 0.98 | 0.84 | 0.67 | 1.00 | 0.93 | 0.83 | 1.00 | 0.98 | 0.93 |
| CV | 1.04 | 1.07 | 1.13 | 1.04 | 1.06 | 1.12 | 1.03 | 1.03 | 1.11 | 1.01 | 1.02 | 1.06 | 1.01 | 1.01 | 1.02 |

Proportion Sold 0.88 0.94 0.98 0.85 0.93 0.97 0.74 0.82 0.90 0.59 0.62 0.67 0.28 0.34 0.38
Table 2: Monte Carlo results: Power and bounds of confidence interval

<table>
<thead>
<tr>
<th>$\tau$ =</th>
<th>-0.8</th>
<th>-0.5</th>
<th>0</th>
<th>0.5</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{CV}$ =</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$N=300$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1.00</td>
<td>0.94</td>
<td>1.00</td>
<td>0.96</td>
<td>0.74</td>
</tr>
<tr>
<td>Lower bound</td>
<td>-1.14</td>
<td>-1.25</td>
<td>-1.27</td>
<td>-0.82</td>
<td>-0.91</td>
</tr>
<tr>
<td>Upper bound</td>
<td>-0.54</td>
<td>-0.35</td>
<td>-0.60</td>
<td>-0.25</td>
<td>-0.13</td>
</tr>
<tr>
<td>$N=1,000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Lower bound</td>
<td>-0.99</td>
<td>-1.04</td>
<td>-1.14</td>
<td>-0.67</td>
<td>-0.72</td>
</tr>
<tr>
<td>Upper bound</td>
<td>-0.65</td>
<td>-0.57</td>
<td>-0.74</td>
<td>-0.37</td>
<td>-0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_s$ =</th>
<th>-0.8</th>
<th>-0.5</th>
<th>0</th>
<th>0.5</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{CV}$ =</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$N=300$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1.00</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Lower bound</td>
<td>0.75</td>
<td>0.55</td>
<td>0.62</td>
<td>0.75</td>
<td>0.62</td>
</tr>
<tr>
<td>Upper bound</td>
<td>1.39</td>
<td>1.59</td>
<td>1.67</td>
<td>1.38</td>
<td>1.53</td>
</tr>
<tr>
<td>$N=1,000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Lower bound</td>
<td>0.87</td>
<td>0.82</td>
<td>0.85</td>
<td>0.89</td>
<td>0.81</td>
</tr>
<tr>
<td>Upper bound</td>
<td>1.22</td>
<td>1.34</td>
<td>1.44</td>
<td>1.20</td>
<td>1.31</td>
</tr>
</tbody>
</table>
Table 3: Bidder summary statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th># Auctions</th>
<th>Percent of total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bidder incumbency status:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Auctions without incumbent</td>
<td>Number of auctions without incumbent or with an incumbent who is not eligible to bid</td>
<td>161</td>
<td>50.79</td>
</tr>
<tr>
<td># Auctions with incumbent &amp; RTM</td>
<td>Number of auctions where player's previous team is eligible to use a RTM card</td>
<td>75</td>
<td>23.66</td>
</tr>
<tr>
<td>of which: # Auctions where RTM used</td>
<td>Number of auctions where player's previous team uses RTM option</td>
<td>13</td>
<td>4.10</td>
</tr>
<tr>
<td># Auctions with incumbent &amp; no RTM</td>
<td>Number of auctions where player's previous team is not eligible to use a RTM card</td>
<td>81</td>
<td>25.55</td>
</tr>
<tr>
<td><strong>Bidder purchases in the auctions:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Players</td>
<td>Number of players bought in the auctions</td>
<td>16.88</td>
<td>4.22</td>
</tr>
<tr>
<td>speciality: # Batsman</td>
<td>Number of batsmen bought in the auctions</td>
<td>4.12</td>
<td>2.58</td>
</tr>
<tr>
<td>speciality: # Bowler</td>
<td>Number of bowlers bought in the auctions</td>
<td>7.25</td>
<td>1.98</td>
</tr>
<tr>
<td>speciality: # Wicket-Keeper</td>
<td>Number of wicket-keepers bought in the auctions</td>
<td>1.62</td>
<td>0.74</td>
</tr>
<tr>
<td>speciality: # All-Rounder</td>
<td>Number of all-rounders bought in the auctions</td>
<td>3.87</td>
<td>1.64</td>
</tr>
<tr>
<td>Nationality: # Indian</td>
<td>Number of Indian players bought in the auctions</td>
<td>11.25</td>
<td>3.01</td>
</tr>
<tr>
<td># Newcomers</td>
<td>Number of players bought in the auctions who are newcomers</td>
<td>7.12</td>
<td>2.10</td>
</tr>
<tr>
<td># RTM Bought Players</td>
<td>Number of players bought in the auctions through RTM</td>
<td>1.62</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Bidder constraints in the auctions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Retained Players</td>
<td>Number of players retained by teams before the auctions</td>
<td>3</td>
<td>1.85</td>
</tr>
<tr>
<td># RTM cards</td>
<td>Number of RTM cards received by teams</td>
<td>1.63</td>
<td>0.74</td>
</tr>
<tr>
<td># of players where incumbent</td>
<td>Number of players eligible to be bought back using RTM (per team) if he has a RTM card</td>
<td>22.37</td>
<td>4.56</td>
</tr>
<tr>
<td>Spending cap</td>
<td>Amount of money allocated to a team</td>
<td>5.65</td>
<td>2.14</td>
</tr>
<tr>
<td>Remaining budget</td>
<td>Unused budget of a team at the end of the auctions</td>
<td>0.60</td>
<td>0.42</td>
</tr>
<tr>
<td># Remaining RTM cards</td>
<td>Unused RTM card of a team at the end of the auctions</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note: All monetary values are reported in millions of USD. The currency used for the 2014 auctions was Indian Rupees (INR); we convert them to USD using an approximate conversion rate of 1 (USD) to 62 (INR).*
Table 4: Summary statistics on auction and performance data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># active bidders</td>
<td>Participating bidders for each player auction</td>
<td>0.87</td>
<td>1.29</td>
</tr>
<tr>
<td>Reserve price</td>
<td>Reservation wage set by auctioneer</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>Order</td>
<td>Within-set order of player appearance in auction</td>
<td>5.23</td>
<td>2.80</td>
</tr>
<tr>
<td>Indian</td>
<td>Dummy indicating whether player is Indian</td>
<td>0.44</td>
<td>0.50</td>
</tr>
<tr>
<td>Newcomer</td>
<td>Dummy indicating whether player is a newcomer</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>Auction with incumbent</td>
<td>Dummy indicating whether one team is an incumbent</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>of which RTM</td>
<td>Dummy indicating whether the incumbent is eligible</td>
<td>0.23</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>to use a RTM card</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sold-players sample:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># active bidders</td>
<td>Participating bidders for each player auction</td>
<td>2.27</td>
<td>1.07</td>
</tr>
<tr>
<td>Winning price</td>
<td>Equal to the final wage of the player</td>
<td>0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Reserve price</td>
<td>Reservation wage set by auctioneer</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Order</td>
<td>Within-set order of player appearance in auction</td>
<td>5.28</td>
<td>2.90</td>
</tr>
<tr>
<td>Indian</td>
<td>Dummy indicating whether player is Indian</td>
<td>0.66</td>
<td>0.48</td>
</tr>
<tr>
<td>Newcomer</td>
<td>Dummy indicating whether player is a newcomer</td>
<td>0.42</td>
<td>0.49</td>
</tr>
<tr>
<td>Auction with incumbent</td>
<td>Dummy indicating whether one team is an incumbent</td>
<td>0.74</td>
<td>0.43</td>
</tr>
<tr>
<td>of which RTM</td>
<td>Dummy indicating whether the incumbent is eligible</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>Winner is incumbent</td>
<td>Dummy indicating whether player was matched with the</td>
<td>0.30</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>winner in the previous season (formally it corresponds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>to $x_{f^\text{dum}}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tournament Points$^*$</td>
<td>Player performance in tournament</td>
<td>23.90</td>
<td>14.55</td>
</tr>
</tbody>
</table>

Note: All monetary values are reported in millions of USD. The currency used for the 2014 auctions was Indian Rupees (INR); we convert them to USD using an approximate conversion rate of 1 (USD) to 62 (INR). $^*$ Construction of the composite performance measure is described in the Appendix;
<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage ($\tau$)</td>
<td>5.80***</td>
<td>4.14**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[2.50, 9.07]</td>
<td>[0.15, 8.11]</td>
<td></td>
</tr>
<tr>
<td><strong>Team/player characteristics</strong> ($x$):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indian</td>
<td>-4.68</td>
<td>-2.52***</td>
<td>-3.22</td>
</tr>
<tr>
<td></td>
<td>[-10.66, 1.30]</td>
<td>[-3.35, -1.68]</td>
<td>[-10.31, 4.48]</td>
</tr>
<tr>
<td>Newcomer</td>
<td>-1.17</td>
<td>0.51</td>
<td>-2.43</td>
</tr>
<tr>
<td></td>
<td>[-9.54, 7.20]</td>
<td>[-0.10, 1.13]</td>
<td>[-11.58, 5.71]</td>
</tr>
<tr>
<td>Speciality: Batsman</td>
<td>2.16</td>
<td>2.16***</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>[-5.76, 10.08]</td>
<td>[1.12, 3.20]</td>
<td>[-3.02, 11.88]</td>
</tr>
<tr>
<td>Speciality: Bowler</td>
<td>8.52**</td>
<td>3.01***</td>
<td>10.15***</td>
</tr>
<tr>
<td></td>
<td>[1.64, 15.39]</td>
<td>[2.01, 4.02]</td>
<td>[3.47, 17.97]</td>
</tr>
<tr>
<td>Incumbent present</td>
<td>1.87</td>
<td>2.57***</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>[-4.57, 8.31]</td>
<td>[1.27, 3.87]</td>
<td>[-3.05, 15.49]</td>
</tr>
<tr>
<td>Bidder (Winner) is incumbent</td>
<td>-2.27</td>
<td>2.23***</td>
<td>-15.27**</td>
</tr>
<tr>
<td></td>
<td>[-8.26, 3.71]</td>
<td>[1.23, 3.21]</td>
<td>[-30.80, -1.47]</td>
</tr>
<tr>
<td><strong>Auction variables</strong> ($z$):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order of sale</td>
<td>-0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.06, 0.05]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remaining budget (in logs)</td>
<td>0.55***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.29, 0.82]</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Batsman bought</td>
<td>-0.36***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.58, -1.14]</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Bowlers bought</td>
<td>-0.48***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.64, -0.32]</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Wicket-keepers bought</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td># All-rounders bought</td>
<td>-0.41***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.24, 0.58]</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Overseas players bought</td>
<td>-0.63***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.80, -0.46]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incumbent present &amp; RTM card not eligible</td>
<td>-1.33***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-2.01, -0.65]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other Structural Parameters:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma^{PV}$</td>
<td>-0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-4.62, 2.78]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma^{CV}$</td>
<td>4.01*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.39, 55.18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.77***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.60, 0.94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^{PV}$</td>
<td>1.48***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.75, 2.92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^{CV}$</td>
<td>2.22***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[1.72, 2.87]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** All specifications account for fixed effects with respect to the set in which the player was auctioned. Column 1 reports OLS estimates of the parameters in the performance equation (1), and 95% CIs based on the usual OLS standard errors. Column 2 gives the ML estimates of the auction model primitives $\beta^*, \beta^r$, $\theta$, $\sigma^{PV}$, and $\sigma^{CV}$, and 95% CIs based on the asymptotic ML standard errors. Column 3 reports OLS estimates of the parameters in the augmented performance equation (10), and 95% CIs based on a percentile bootstrapped procedure (with 1,000 bootstrapped samples). * indicates significance at 10%; ** at 5%; *** at 1%.
A Appendix

A.1 Proof of Proposition 3.1

We show here why the bidding strategy $[\hat{bs}^C_f]^{-1}(.)$ corresponds to firm $f$’s best response in case C and taking as given the incumbent bidding strategy $[\hat{bs}^C_{inc}]^{-1}(.)$. Let $\Pi^C_f(bs_f, w)$ denote firm $f$’s expected payoff if his bid-signal is $bs_f$, if he remains active until price $w$ and if the incumbent has entered the auction and is still active while bidding optimally when the incumbent is no longer active. Let $\Pi^B_f(bs_f, w)$ denote the continuation payoff of firm $f$ if the incumbent has dropped out at price $w$, if there is another non-incumbent which is active a price $w$ and if he bids optimally afterwards (e.g. by dropping out immediately after the incumbent’s exit). Let $p_f(w)$ denote the belief of firm $f$ about the probability that the auction stops if the incumbent drops out at price $w$. Thanks to our full support assumption on bid-signals, we have $p_f(w) > 0$ for any $w > 0$. As we will see, the exact value $p_f(.)$ (which depends on the other non-incumbents’ strategies) does not play any role to characterize firm $f$ optimal bidding strategy. We have that

$$\Pi^C_f(bs_f, b) = \int_{\hat{bs}^C_{inc}(W^*)}^{\hat{bs}^C_{inc}(b)} \Pi_f(bs_f, [\hat{bs}^C_{inc}]^{-1}(x)) \cdot \frac{dG_{inc}(x)}{1 - G_{inc}(bs^C_{inc}(W^*))}$$

where

$$\Pi_f(bs_f, w) := [w^\theta \cdot e^{(1-\theta)[A_{inc}+bs_f+U_{inc}(\hat{bs}^C_{inc}(w))]} - w] \cdot p_f(w) + \Pi^B_f(bs_f, w) \cdot (1 - p_f(w)).$$

By definition of $\hat{bs}^C_f(.)$, we have $[[\hat{bs}^C_f]^{-1}(bs_f)]^\theta \cdot e^{(1-\theta)[A_{inc}+bs_f+U_{inc}(\hat{bs}^C_{inc}([\hat{bs}^C_f]^{-1}(bs_f)))]} = [\hat{bs}^C_f]^{-1}(bs_f)$. Furthermore, if the incumbent drops exactly at $[\hat{bs}^C_f]^{-1}(bs_f)$, then firm $f$’s best strategy is then to drop out immediately which yields a null payoff. We have thus $\Pi^B_f(bs_f, [\hat{bs}^C_f]^{-1}(bs_f)) = 0$. On the whole we have $\Pi_f(bs_f, [\hat{bs}^C_f]^{-1}(bs_f)) = 0$.

In order to show that the optimal bid is $[\hat{bs}^C_f]^{-1}(bs_f)$, we establish next that $\Pi_f(bs_f, w) > (\text{resp.} <) 0$ if $w < (\text{resp.} >) [\hat{bs}^C_f]^{-1}(bs_f)$ for any possible strictly positive value for $p_f(w)$. Since $\hat{bs}^C_f(.)$ is increasing, we get that $w^\theta \cdot e^{(1-\theta)[A_{inc}+bs_f+U_{inc}(\hat{bs}^C_{inc}(w))]} - w > (\text{resp.} <) 0$ if $w < (\text{resp.} >) [\hat{bs}^C_f]^{-1}(bs_f)$.

For $w < [\hat{bs}^C_f]^{-1}(bs_f)$, we have $\Pi^B_f(bs_f, w) \geq 0$: firm $f$ can always guarantee himself a positive payoff by dropping out immediately at the price $w$. On the contrary, for any $w > [\hat{bs}^C_f]^{-1}(bs_f)$, we have $\Pi^B_f(bs_f, w) \leq 0$: firm $f$ would like to drop out immediately and would raise a negative payoff if he wins the auction at price $w$. Q.E.D.
A.2 Details on the control functions

Case 3:

From Proposition 3.1 we know that \( w_i = W_i^r \) and \( f_i^w \neq f_i^{inc} \) if and only if \( bs_{i,f_i} \leq \tilde{b}_s_{i,f_i}(W_i^r) \), \( bs_{i,f_i} \geq \tilde{b}_s_{i,f_i}(W_i^r) \) and \( bs_{i,f} \leq \tilde{b}_s_{i,f}(W_i^r) \) for each \( f \neq f_i^{inc}, f_i^w \). Given A1, we have that conditional on \( \mathcal{F} \), \( bs_{i,f_i} \) (resp. \( bs_{i,f_i}^{inc} \)) is distributed according to the distribution \( G_{f_i}^{PV} \) (resp. \( G_{f_i}^{inc} \)) truncated below \( \tilde{b}_s_{i,f_i}(W_i^r) \) (resp. above \( \tilde{b}_s_{i,f_i}^{inc}(W_i^r) \)). Using the same arguments as in cases 1 and 2, we obtain

\[
CF_{i,f_i}^{PV}[1] = \int_{-\infty}^{+\infty} x^l \cdot \frac{dG_{f_i}^{PV}(x)}{1 - G_{f_i}^{PV}(\tilde{b}_s_{i,f_i}(W_i^r))}
\]

and

\[
CF_{i}^{CV}[1] = \int_{-\infty}^{\tilde{b}_s_{i,f_i}(W_i^r)} E_{x \sim G_{f_i}^{CV}[\tilde{x}, f_i^{inc}]}[x^l] \frac{dG_{f_i}^{inc}(\tilde{x})}{G_{f_i}^{inc}(\tilde{b}_s_{i,f_i}(W_i^r))} \cdot d\tilde{x}.
\]

Under the symmetric Gaussian structure, we get the Mills ratios:

\[
CF_{i,f_i}^{PV}[1] = \sigma_{PV} \cdot \frac{\phi(\tilde{b}_s_{i,f_i}(W_i^r))}{1 - \phi(\tilde{b}_s_{i,f_i}(W_i^r))} \quad \text{and} \quad CF_{i}^{CV}[1] = -\frac{\sigma_{CV}^2}{\sqrt{\sigma_{PV}^2 + \sigma_{CV}^2}} \cdot \frac{\phi(\tilde{b}_s_{i,f_i}(W_i^r))}{\phi(\tilde{b}_s_{i,f_i}(W_i^r))}.
\]

To get (16), we develop the expressions of \( \tilde{b}_s_{i,f_i}(W_i^r) \) and \( \tilde{b}_s_{i,f_i}^{inc}(W_i^r) \) in (refCFcase3Bis). In particular, we develop the third term appearing in \( \tilde{b}_s_{i,f_i}(W_i^r) \), namely

\[
U_i^A = \frac{1}{(1 - \theta)} \cdot \log(E[e^{(1-\theta)(b_{1}^{inc}+b_{2}^{inc})}|bs_{i,f_i}^{inc} \leq 1 - \theta \cdot (1 - \theta) \cdot (1 - \theta) \cdot \sigma_{2}^2 \cdot a \quad \text{and variance} \quad (1 - \rho^2) \cdot \sigma_{2}^2]. \quad \text{We obtain then that} \quad E[e^{\gamma_2}|y_2 = a] = \exp(\rho \cdot \frac{a}{\sigma_{2}^2} \cdot a + (1 - \rho^2) \cdot \frac{\sigma_{2}^2}{2}) \quad \text{and then that} \quad E[e^{\gamma_2}|y_2 \leq a] = E[\exp(\rho \cdot \frac{a}{\sigma_{2}^2} \cdot y_2)|y_2 \leq a] \cdot \exp((1 - \rho^2) \cdot \frac{\sigma_{2}^2}{2}). \quad \text{Then using the formula of the expectation of a truncated normal distribution (see chapter 24 in Greene (2008)), we}
\]

53
obtain finally that

\[ E[e^{y_1}|y_2 \leq a] = e^{\frac{a}{\sigma_2^2}} \cdot \frac{1 - \Phi(\rho \cdot \sigma_1 - \frac{a}{\sigma_2})}{\Phi(\frac{a}{\sigma_2})}. \]

Applying this formula to the jointly normally distributed vector \(((1 - \theta) \cdot (bs_i^{co} + bs_i^{CV}, bs_i^{inc}))\), and after some simple algebra, we get indeed

\[ U_i^A = \frac{1}{1 - \theta} \log \left( \frac{1 - \Phi(\frac{(1-\theta)\sigma_{pv}^{2} - \log(W) - A_{i,inc}}{\sqrt{\sigma_{pv}^{2} + \sigma_{cv}^{2}}})}{\Phi(\frac{\log(W) - A_{i,inc}}{\sqrt{\sigma_{pv}^{2} + \sigma_{cv}^{2}}})} \right) + \frac{(1-\theta)}{2} \cdot \sigma_{cv}^{2}. \]

**Case 4:**

For \( l = 1, \ldots, L \), let \( CF_{i,f}^{PV}(f)[l] := E[(bs_{i,f}^{id} + bs_{i,f}^{CV}) | f_i^{sh} = f \] and \( CF_{i,f}^{CV}(f)[l] := E[(bs_{i}^{co} + bs_{i}^{CV}) | f_i^{sh} = f \] denote the analogs of control functions defined in Section 4, except that here we condition on the identity of the second-highest, denoted by \( f_i^{sh} \). Recall that the identity of this bidder is not assumed to be observed by the econometrician. To calculate the conditional control functions \( CF_{i,f}^{PV}(f)[l] \) and \( CF_{i,f}^{CV}(f)[l] \), we distinguish two sub-cases: 4a) the incumbent is the second-highest bidder; 4b) the incumbent is not the second-highest bidder.

**Case 4a** \( (f_i^{sh} = f_i^{inc}) \): From Proposition 3.1 we know that \( f_i^{sh} = f_i^{inc} \) if and only if \( bs_{i,inc} = \tilde{bs}_{i,inc}(w_i) \), \( bs_{i,inc} \geq \tilde{bs}_{i,f}^{C}(w_i) \) and \( bs_{i,f} \leq \tilde{bs}_{i,f}^{C}(w_i) \) for each \( f \neq f_i^{inc}, f_i^{w} \). Given A1 and using similar arguments as for the first three cases, we obtain

\[ CF_{i,f}^{PV}(f^{inc})[l] = \int_{bs_{i,f}^{C}(w_i)}^{+\infty} \frac{dG_{f_i^{inc}}^{PV}(x)}{1 - G_{f_i^{inc}}^{PV}(bs_{i,f}^{C}(w_i))} \quad \text{and} \quad CF_{i,f}^{CV}(f^{inc})[l] = \int_{-\infty}^{\tilde{bs}_{i,inc}(w_i)} \frac{g_{f_i^{inc}}^{CV}(bs_{i,inc}(w_i) - x)}{g_{f_i^{inc}}^{PV}(bs_{i,inc}(w_i))} dx. \]

Under the symmetric Gaussian structure, we get tractable expressions for the cutoff bid-signals \( \tilde{bs}_{i,inc}(w_i) \) and \( bs_{i,inc}(w_i) \) (the former is defined in Section 3.3 and the latter in (3)): \( \tilde{bs}_{i,inc}(w_i) = \log(w_i) - A_{i,inc} \) and \( bs_{i,inc}(w_i) = \log(w_i) - A_{i,inc} - U_{i,inc}(\tilde{bs}_{i,inc}(w_i)) \) with

\[ U_{i,inc}(x) = U(x) = \frac{\sigma_{CV}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \cdot \left( x + \frac{1-\theta}{2} \cdot \sigma_{pv}^{2} \right). \]

A formula which comes from the same computations as for case 3 where we compute \( U_{i,inc}^{A} \).

Finally, for \( l = 1 \), we get the expression

\[ CF_{i,f}^{PV}(f^{inc})[l] = \sigma_{pv} \cdot \left( \frac{\sigma_{pv}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \cdot \log(w_i) - \frac{1}{\sigma_{pv}} \cdot A_{i,f}^{w} + \frac{\sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \cdot A_{i,f}^{inc} - \frac{1-\theta}{2} \cdot \frac{\sigma_{pv} \cdot \sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \right) \]

\[ - \Phi \left( \frac{\sigma_{pv}^{2} \cdot \log(w_i) - \frac{1}{\sigma_{pv}} \cdot A_{i,f}^{w} + \frac{\sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \cdot A_{i,f}^{inc} - \frac{1-\theta}{2} \cdot \frac{\sigma_{pv} \cdot \sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}}}{\sqrt{\sigma_{pv}^{2} + \sigma_{cv}^{2}}} \right) \]

\[ - \Phi \left( \frac{\sigma_{pv}^{2} \cdot \log(w_i) - \frac{1}{\sigma_{pv}} \cdot A_{i,f}^{w} + \frac{\sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}} \cdot A_{i,f}^{inc} - \frac{1-\theta}{2} \cdot \frac{\sigma_{pv} \cdot \sigma_{cv}^{2}}{\sigma_{pv}^{2} + \sigma_{cv}^{2}}}{\sqrt{\sigma_{pv}^{2} + \sigma_{cv}^{2}}} \right). \]
Analogously as in case 3, the control function $C_{i_i}^{PV}(f_i^{inc})$ depends on the variables $x_{i,f}$ and $z_{i,f}$, for both $f = f_i^w$ and $f = f_i^{inc}$.

**Case 4b** $f_i^{sh} \neq f_i^{inc}$: This case occurs if either the incumbent did not enter the auction at all, or the incumbent has participated but quit the auction before the second highest bidder. Let $\pi_{i_{f_{inc}}}^{IP}(w_i, f_i^w, f_i^{sh})$ denote the probability that the incumbent has not entered the auction for worker $i$ conditional on the observable auction outcome $(w_i, f_i^w, f_i^{sh})$. Let $\pi_{i_{f_{inc}}}^{P}(p|w_i, f_i^w, f_i^{sh})$ denote the CDF of the drop-out price of the incumbent conditional on the auction outcome $(w_i, f_i^w, f_i^{sh})$ and conditional on the incumbent having entered the auction. Note that the support of the distribution $\pi_{i_{f_{inc}}}^{P}(.|w_i, f_i^w, f_i^{sh})$ is $[W_i^r, w_i]$. Equipped with these notation, given $A_1$ and by iterated expectations, we now have in case 4b:

$$CF_{i_{f_i}^{w}}^{PV}(f_i^{sh})[l] = \pi_{i_{f_{inc}}}^{IP}(w_i, f_i^w, f_i^{sh}) \cdot \int_{bs_{i_{f_i}^{w}}(w_i)}^{\infty} \frac{d[G_{i_{f_i}^{w}}^{PV}(x)]}{1 - G_{i_{f_i}^{w}}^{PV}(bs_{i_{f_i}^{w}}(w_i))} x^l \cdot \frac{d[G_{i_{f_i}^{w}}^{PV}(x)]}{1 - G_{i_{f_i}^{w}}^{PV}(bs_{i_{f_i}^{w}}(w_i, p))} \cdot d\pi_{i_{f_{inc}}}^{P}(p|w_i, f_i^w, f_i^{sh})$$

and

$$CF_{i_{f_i}^{w}}^{CV}(f_i^{sh})[l] = \pi_{i_{f_{inc}}}^{IP}(w_i, f_i^w, f_i^{sh}) \cdot CF[3] + (1 - \pi_{i_{f_{inc}}}^{IP}(w_i, f_i^w, f_i^{sh})) \cdot \int_{W_i^r}^{w_i} CF[4b; p] \cdot d\pi_{i_{f_{inc}}}^{P}(p|w_i, f_i^w, f_i^{sh})$$

where $CF[3]$ corresponds to the expression of the control function $CF_{i_{f_i}^{w}}^{CV}[l]$ in case 3, and $CF[4a; p]$ corresponds to the expression of the control function $CF_{i_{f_i}^{w}}^{CV}[l]$ in case 4a but if the final price $w_i$ is replace by $p$ (namely, we have $CF[4a; p] = \frac{\sigma^2_{CV}}{\sigma^2_{PV} + \sigma^2_{CV}} \cdot [\log(p) - A_{i_{f_i}^{inc}}]$).

Letting $p_i^{sh}(\mathcal{S}, f)$ denote the probability that the identity of the second-highest bidder is $f$ conditional on $\mathcal{S}$, we can now, by iterated expectation, relate the unconditional and conditional control functions:

$$CF_{i_{f_i}^{w}}^{PV}[l] = \sum_{f \in \mathcal{S} \setminus \{w_i\}} p_i^{sh}(\mathcal{S}, f) \cdot CF_{i_{f_i}^{w}}^{PV}(f)[l] \quad \text{and} \quad CF_{i_{f_i}^{w}}^{CV}[l] = \sum_{f \in \mathcal{S} \setminus \{w_i\}} p_i^{sh}(\mathcal{S}, f) \cdot CF_{i_{f_i}^{w}}^{CV}(f)[l].$$

Let us now calculate the probability $p_i^{sh}(\mathcal{S}, f)$ for any $f \neq f_i^{inc}$. For $w > W_i^r$, we let $\pi_{i}^{r}(w, f_i^w, f_i^{sh})$ denote the density (from an ex-ante perspective) associated to the event that the auction for worker $i$ stops at wage $w \geq W_i^r$ and that the highest and second-highest bidders
are respectively \( f^w \) and \( f^{sh} \).

If \( f_i^{inc} = f^{sh} \), then

\[
\pi_i^*(w, f^w, f_i^{inc}) = \left[ \prod_{f \neq f_i^{inc}} G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w)) \right] \cdot G_{f_i^{inc}}^{PV}(\tilde{b}_{s_{i,f_i}}(w)) \cdot (1 - G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w))).
\]

If \( f_i^{inc} \neq f^w, f^{sh} \), then either the incumbent did not enter the auction at all, or the incumbent quits the auction before the second-highest bidder. From Proposition 3.1, the first possibility occurs if and only if \( \tilde{b}_{s_{i,f_i}} \leq \tilde{b}_{s_{i,f_i^{inc}}} \) and \( \max_{f \neq f_i^{inc}, f^w} (\tilde{b}_{s_{i,f_i}} \cdot \tilde{b}_{s_{i,f_i^{inc}}}) = w \); the second possibility occurs if and only if there exists a price \( p \in [W_i^r, w] \) at which the incumbent dropped out (before firm \( f_i^{sh} \)), and \( \tilde{b}_{s_{i,f_i^{inc}}} = \tilde{b}_{s_{i,f_i}} \leq \tilde{b}_{s_{i,f_i^{inc}}} \) for any \( f \neq \{f_i^{inc}, f_i^w, f_i^{sh}\} \).

\[
\pi_i^*(w, f^w, f^{sh}) = G_{f_i^{inc}}^{PV}(\tilde{b}_{s_{i,f_i}}(W_i^r)) \cdot \left[ \prod_{f \neq f_i^{inc}} G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w)) \right] \cdot G_{f_i^{inc}}^{PV}(\tilde{b}_{s_{i,f_i}}(w)) \cdot (1 - G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w)))
\]

\[+ \int_{W_i^r}^w \left[ \prod_{f \neq f_i^{inc}} G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w, p')) \right] \cdot G_{f_i^{inc}}^{PV}(\tilde{b}_{s_{i,f_i}}(w, p')) \cdot (1 - G_f^{PV}(\tilde{b}_{s_{i,f_i}}(w, p'))) \cdot d[\tilde{b}_{s_{i,f_i}}(p')] \]

From Bayesian updating (if \( \sum_{f' \neq f_i^{inc}} \pi_i^*(w, f_i^{inc}, f') \neq 0 \), we get finally that

\[
\pi_i^{sh}(s, f) = \frac{\pi_i^*(w_i, f_i^{inc}, f_i^{sh})}{\sum_{f' \neq f_i^{inc}} \pi_i^*(w_i, f_i^{inc}, f')}.
\]

2) Computation of \( \pi_{i,inc}^{NP}(w, f^w, f^{sh}, i) \) and \( \pi_{i,inc}^{NP}(\cdot|w, f^w, f^{sh}, i) \). Let \( g_i^* \) denote the joint density of the vector \( \tilde{b}_s = (\tilde{b}_{s_1}, \ldots, \tilde{b}_{s_F}) \) in auction for worker \( i \). Given independence, \( g_i^*(\tilde{b}_s) = \prod_{f \neq f_i^{inc}} G_f^{PV}(\tilde{b}_{s_f}) \cdot G_{f_i^{inc}}^{inc}(\tilde{b}_{s_f}) \). Let \( \mathcal{S}(w, f^w, f^{sh}, i) \) denote the set of vectors \( \tilde{b}_s \) in \( R^F \) that lead to the auction outcome \( w, f^w, f^{sh} \) in the auction for worker \( i \). Let \( \mathcal{S}(\tilde{b}_s = x) = \{ \tilde{b}_s \in R^F | \tilde{b}_s = x \} \) and \( \mathcal{S}(\tilde{b}_s \leq x) = \{ \tilde{b}_s \in R^F | \tilde{b}_s \leq x \} \). Next we use the notation \( \int_S g_i^*(\tilde{b}_s) d\tilde{b}_s \) (with \( S \) being a subset of \( R^F \)) to denote the density associated to the realization of the event \( S \). Bayesian updating corresponds to:

\[
\pi_{i,inc}^{NP}(w, f^w, f^{sh}, i) = \frac{\int_{\mathcal{S}(w, f^w, f^{sh}, i) \cap \mathcal{S}(\tilde{b}_s \leq \tilde{b}_{s_{i,f_i^{inc}}} (W_i^r))} g_i^*(\tilde{b}_s) \tilde{b}_s \cdot \int_{\mathcal{S}(w, f^w, f^{sh}, i) \cap \mathcal{S}(\tilde{b}_s \leq \tilde{b}_{s_{i,f_i^{inc}}} (W_i^r))} g_i^*(\tilde{b}_s) d\tilde{b}_s}{\int_{\mathcal{S}(w, f^w, f^{sh}, i)} g_i^*(\tilde{b}_s) d\tilde{b}_s}
\]

56
and

\[ \pi^*_f(p|w, f^w, f^{\text{inh}}, f) = \frac{\int_{\mathcal{S}(f, w, f^{\text{inh}}, f^w)} g^*_s(dbs) \, dbs}{(1 - \pi_{\text{inc}}(W, f, w, f^{\text{inh}})) \cdot \int_{\mathcal{S}(f, w, f^{\text{inh}}, f^w)} g^*_s(dbs) \, dbs} \]

Note that the incumbent does not participate if and only if \( \tilde{b}_{s, f, \text{inc}} \leq \tilde{b}_{s, f, \text{inc}}(W) \) and that the incumbent quits the auction at price \( p \) if and only if \( \tilde{b}_{s, f, \text{inc}} \leq \tilde{b}_{s, f, \text{inc}}(p) \).

To complete the computation, note that we have:

\[
\int_{\mathcal{S}(f, w, f^{\text{inh}}, f^w)} g^*_s(dbs) \, dbs = \left[ \prod_{f \in \mathcal{F}^{\text{inh}}, f^{\text{inh}}} \mathcal{G}^A_{f, s}(s^{A}) \right] \cdot \mathcal{G}^A_{f, s}(s(w)) \cdot \mathcal{G}^\text{inc}(s(w)) \cdot \frac{dG^\text{inc}(s(w))}{dp},
\]

and

\[
\int_{\mathcal{S}(f, w, f^{\text{inh}}, f^w)} g^*_s(dbs) \, dbs = \left[ \prod_{f \in \mathcal{F}^{\text{inh}}, f^{\text{inh}}} \mathcal{G}^B_{f, s}(s^{B}) \right] \cdot \mathcal{G}^B_{f, s}(s(w)) \cdot \mathcal{G}^\text{inc}(s(w)) \cdot \frac{dG^\text{inc}(s(w))}{dp},
\]

\[
\int_{\mathcal{S}(f, w, f^{\text{inh}}, f^w)} g^*_s(dbs) \, dbs = \left[ \prod_{f \in \mathcal{F}^{\text{inh}}, f^{\text{inh}}} \mathcal{G}^\text{inc}(s(w)) \cdot \frac{dG^\text{inc}(s(w))}{dp} \right] + \int_{W} \left[ \prod_{f \in \mathcal{F}^{\text{inh}}, f^{\text{inh}}} \mathcal{G}^\text{inc}(s(w)) \cdot \frac{dG^\text{inc}(s(w))}{dp} \right].
\]

A.3 Format of the tournament and player performance measure

In the IPL, a match is generally completed in 3 hours. The match involves one team batting (striking the ball) while the opposing team bowls (delivers the ball), followed by the opposing team batting. The objective of the batting team is to post the maximum amount of score in a certain period of time by striking the ball. The team that posts the highest score wins the match. A batsman is a player who specializes in hitting or ‘striking’ the cricket ball in order to score runs. A bowler is a player who specializes in delivering the ball to a batsman and whose primary aim is to dismiss the batsman or concede minimal runs. A wicket-keeper is a batsman who holds a special position in the field; his role is to stand behind the batsmen and guard the ‘wicket’ when a team is bowling, similar to the role of a catcher in baseball. All-rounders are players who are specialized in, both, batting and bowling. The general composition of a cricket team is three specialist batsmen, four all-rounders, three specialist bowlers and a wicket-keeper. The player specialities are an important feature of our auction model, because teams are implicitly constrained to select and bid in a way that optimizes their team composition (i.e., they are unlikely to buy only bowlers).

Our composite performance measure is derived from various, batting and bowling statistics.
observed for each player during the tournament. The first step in that process was to award points for each basic statistic accumulated by each player across every match of the tournament. The mapping of game-specific player statistics to points is given in Table 6.

Table 6: Conversion of performance statistics into points

<table>
<thead>
<tr>
<th>Performance Statistic</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td># Runs - a batsman's score from striking the ball in the tournament</td>
<td>1 base points for each run</td>
</tr>
<tr>
<td># 50s - Number of times a batsman ended the match with a score equal to or above 50 in the tournament</td>
<td>25 bonus points for each 50</td>
</tr>
<tr>
<td># 100s - Number of times a batsman ended the match with a score equal to or above 100 in the tournament</td>
<td>50 bonus points for each 100</td>
</tr>
<tr>
<td># Wickets - Number of batsman dismissed by a bowler in the tournament</td>
<td>25 base point for each wicket</td>
</tr>
<tr>
<td>4 Wickets - Number of times in the tournament when a bowler dismissed 4 batsman in one match</td>
<td>40 bonus points for 4-wicket haul</td>
</tr>
<tr>
<td>5 Wickets - Number of times in the tournament when a bowler dismissed 5 batsman in one match</td>
<td>50 bonus points for 5-wicket haul</td>
</tr>
</tbody>
</table>

Next, given the time constraints inherent in the format of the game and its emphasis on the rate of scoring, the player's total number of points was adjusted by a speciality specific factor. For batsmen, the factor measures the batsman's relative strike-rate in the tournament, as the higher the strike rate, the more effective a batsman is at scoring quickly. For bowlers, the factor measures the bowler's relative economy-rate in the tournament, as the higher the economy rate, the more effective a bowler is at limiting the opposition's total score. Finally, each player's adjusted points are divided by the number of games they played in the tournament as batsman or bowlers so that player's are judged on a per-game basis.

---

56See https://bit.ly/2CvCB44 for a description of an algorithm that constructs a similar performance measure to compare and rank players in the tournament. Note that the performance measure distinguishes only batsmen and bowlers; since wicket-keepers and all-rounders either bat (wicket-keepers) or bat and bowl (all-rounders), their performance is accounted through their batting and/or bowling scores.

57A batsman's strike-rate is defined as the average score of a batsman per 100 balls faced. Formally, this is equal to \[100 \times (\text{Batsman score} / \# \text{ Balls faced})\]. The batsman's factor is his strike-rate divided by the average strike-rate of other batsmen in the tournament.

58A bowler's economy-rate is defined as the average score conceded by a bowler per 6 balls. Formally, this is equal to \[\text{Bowler Score} / (\# \text{ Balls delivered} / 6)\]. The bowler's factor is his economy-rate divided by the average economy-rate of other bowlers in the tournament.
Appendix (NOT FOR PUBLICATION)

B.1 Deconvolution formula

If we let \( \tilde{H} \) denote the Fourier transform of the CDF \( H \), namely we have a bijection between \( H \) and \( \tilde{H} \) characterized by the relations \( \tilde{H}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} H(x)e^{-i\xi x}dx \) and \( H(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \tilde{H}(\xi)e^{+i\xi x}d\xi \). The Fourier transform of the CDF of the sum of two independent variables is the product of the Fourier transform of the two underlying variables. \( A6 \) implies then that \( \tilde{G}_{f \text{inc}}(\xi) = \tilde{G}^{PV}_{f \text{inc}}(\xi) \cdot \tilde{G}^{CV}(\xi) \). Finally \( G^{CV} \) is formally characterized as a function of the CDF \( G_{f \text{inc}} \) and \( G^{PV}_{f \text{inc}} \) by \( G^{CV}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G_{f \text{inc}}(x')e^{-i\xi x'}dx' \cdot e^{+i\xi x}d\xi \).

B.2 Likelihood function

Here we derive the likelihood function of the auction data when there is no RTM. The likelihood depends on the parameters \( \theta, \alpha, \beta^*, \beta'^* \), \( f = 1, \cdots, F \), and the parameters characterizing the (parametrized) distribution functions \( G^{CV}, G^{PV}_f, \) and \( G_f \). Note that the cutoff bid-signals \( \tilde{b}^{s_f}_k \) and \( \tilde{b}^{A}_k \), \( k = A, B, C \), depend on all these parameters. For notational simplicity we will not explicitly index the cutoff bid-signals, the distribution functions, and the other entities defined below, by the parameter vector. We will only give the likelihood function for observations where an incumbent is present among the potential bidders. The likelihood for auctions without an incumbent corresponds to the likelihood of a standard independent PV English auction. Its form can be found in for example Baldwin, Marshall, and Richard (1997).

Let \( P_{i,f \text{inc}}(w) \) denote the ex-ante probability that the incumbent prefers not to employ worker \( i \) at wage \( w \) given the value of the parameters. We thus have \( P_{i,f \text{inc}}(w) = G_{f \text{inc}}(\tilde{b}^{s_i,f \text{inc}}(w)) \), where \( \tilde{b}^{s_i,f \text{inc}}(w) \) is defined in (3), and \( p_{i,f \text{inc}}(w) = g_{f \text{inc}}(\tilde{b}^{s_i,f \text{inc}}(w)) \) is the corresponding density evaluated at the cutoff bid-signal of the incumbent. Using the letters \( A, B, \) and \( C \) associated with the three cases described in Section 3, we similarly define \( P^A_{i,f}(w) \) (resp. \( P^B_{i,f}(w, p) \) for \( p \in [W_i^r, w] \)) as the probability that firm \( f (f \neq f \text{inc}) \) prefers not to employ \( i \) at wage \( w \), conditional on observing that the incumbent has not entered the auction (resp. has entered and dropped out at price \( p \)). We have thus \( P^A_{i,f}(w) = G^PV_f(\tilde{b}^{A}_{i,f}(w)) \) and \( P^B_{i,f}(w, p) = G^PV_f(\tilde{b}^{B}_{i,f}(w, p)) \), where the cutoff bid-signals \( \tilde{b}^{A}_{i,f}(w) \) and \( \tilde{b}^{B}_{i,f}(w, p) \) are defined in Section 3. We also define \( P^C_{i,f}(w) \) as the probability that firm \( f \) does not wish to employ \( i \) at \( w \), given that the incumbent is still active at \( w \) and that this firm believes the incumbent is going to quit instantly at this price. This probability can hence be written as \( P^C_{i,f}(w) = G^PV_f(\tilde{b}^{C}_{i,f}(w)) \). Finally, let \( p^k_{i,f}(w) = g^PV_f(\tilde{b}^{k}_{i,f}(w)) \) for \( k = A, C \) and \( p^B_{i,f}(w, p) = g^PV_f(\tilde{b}^{B}_{i,f}(w, p)) \), the densities evaluated at the appropriate cutoff bid-signals.

For a given value of the parameter vector, the likelihood associated with the event that worker
\( i \) remains unsold conditional on \( \mathcal{S} \), is denoted \( l_{i}^{\text{unsold}} \). Given Proposition 3.1 and the independence assumption A1, we have:

\[
L_{i}^{\text{unsold}} = \prod_{f=1}^{F} P_{i,f}^{C}(W_{i}^{r}) \times P_{i,f_{i}^{\text{inc}}}(W_{i}^{r}).
\]

To write down the other terms of the likelihood function, we now use that the identity of the second-highest bidder is observed by the econometrician. Letting \( f_{i}^{sh} \) be the identity of the the second-highest bidder in auction \( i \) (if there is not one, then \( f_{i}^{sh} := 0 \)), it is thus assumed that \( \mathcal{S} \) contains \( f_{i}^{sh} \) for all \( i \). The likelihood associated with the event that \( i \) is sold to firm \( f_{i}^{w} \) at \( w_{i} \), and the second highest bidder is \( f_{i}^{sh} \), conditional on \( \mathcal{S} \), is denoted \( L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) \). The precise form of this type of likelihood contribution depends on whether the incumbent is the winner, the second highest bidder, or neither of these two bidders. It also depends on whether \( i \) is sold at or strictly above the reserve price.

If there is a single entrant (so that \( w_{i} = W_{i}^{r} \) and \( f_{i}^{sh} = 0 \)), we have:

\[
L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) = \prod_{f=1}^{F} P_{i,f}^{C}(W_{i}^{r}) \times \left( 1 - P_{i,f_{i}^{w}}(W_{i}^{r}) \right) \text{ if } f_{i}^{w} = f_{i}^{\text{inc}},
\]

and

\[
L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) = P_{i,f_{i}^{\text{inc}}}(W_{i}^{r}) \times \prod_{f=1}^{F} P_{i,f}^{C}(W_{i}^{r}) \times \left( 1 - P_{i,f_{i}^{w}}(W_{i}^{r}) \right) \text{ if } f_{i}^{w} \neq f_{i}^{\text{inc}}.
\]

If there are at least two entrants (so that \( w_{i} > W_{i}^{r} \) and \( f_{i}^{sh} \neq 0 \)), we have:

\[
L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) = \prod_{f=1}^{F} P_{i,f}^{C}(w_{i}) \times P_{i,f_{i}^{\text{inc}}}(w_{i}) \times \left( 1 - P_{i,f_{i}^{w}}(w_{i}) \right) \text{ if } f_{i}^{w} = f_{i}^{\text{inc}},
\]

\[
L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) = \prod_{f=1}^{F} P_{i,f}^{C}(w_{i}) \times P_{i,f_{i}^{\text{inc}}}(w_{i}) \times \left( 1 - P_{i,f_{i}^{w}}(w_{i}) \right) \text{ if } f_{i}^{w} \neq f_{i}^{\text{inc}}.
\]

and

\[
L_{i}^{\text{sold}}(w_{i}, f_{i}^{w}, f_{i}^{sh}) = \prod_{f=1}^{F} P_{i,f}^{A}(w_{i}) \times P_{i,f_{i}^{\text{inc}}}(w_{i}) \times \left( 1 - P_{i,f_{i}^{w}}(w_{i}) \right) \times P_{i,f_{i}^{\text{inc}}}(W_{i}^{r})
\]

60
B.3 Warp-Speed Monte Carlo

The Warp-Speed method will be described by considering the parameter \( \tau \). The methodology is strictly the same for any other parameter. From the \( k \)-th replication sample (\( k = 1, \cdots, 1,000 \)), we draw a single bootstrap sample of size \( N \). Letting \( \hat{\tau}_{N,k} \) and \( \hat{\tau}^*_{N,k} \) be the second-stage estimates using the \( k \)-th Monte Carlo sample and its associated bootstrap resample, respectively, we can then construct a sequence of 95\% confidence intervals for \( \tau \)

\[
CI_{N,k}(\tau) = [\hat{\tau}_{N,k} - q_N(0.975), \hat{\tau}_{N,k} - q_N(0.025)], \quad \text{and} \quad k = 1, \cdots, 1,000,
\]

where \( q_N(0.025) \) and \( q_N(0.975) \) are the 0.025-quantile and 0.975-quantile of the empirical distribution of \( \hat{\tau}^*_{N,k} - \hat{\tau}_{N,k}, k = 1, \cdots, 1,000 \), respectively. We can now estimate the lower bound (resp. upper bound) of the 95\% confidence interval of \( \tau \) by taking the mean over the lower bounds (resp. upper bounds) of \( CI_{N,k}(\tau), k = 1, \cdots, 1,000 \). Similarly, the power of the t-test of the null hypothesis \( H_0: \tau = 0 \) (against the bilateral alternative) can simply be estimated by the fraction of times zero does not belong to \( CI_{N,k}(\tau), k = 1, \cdots, 1,000 \), given that the data are generated under a particular value of \( \tau \). The novelty of the method proposed by Giacomini, Politis, and White (2013) is that only one bootstrap resample is required for each replication (instead of some high number as in a standard Monte Carlo experiment), thereby drastically reducing the computation time.