

Correcting for Sample Selection From Competitive Bidding, with an Application to Estimating the Effect of Wages on Performance*

Laurent Lamy[†]

Manasa Patnam[‡]

Michael Visser[§]

July 6, 2016

Abstract

This paper proposes a method to estimate the relationship between the price of a good sold at auction, and a post-auction outcome which is observed among auction winners. To account for both the endogeneity of the auction price and sample selection, we develop a control function approach based on the non-parametric identification of an auction model. In our application we estimate a performance equation using unique field data on wages earned by cricket players and their game-specific performances. Our empirical strategy benefits from the fact that wages are determined through randomly ordered sequential English auctions: the order in which players are sold acts as an exogenous shifter of wages. We find that the positive correlation between wages and performance comes (almost) exclusively from the selection and endogeneity effects.

Keywords: Sample Selection; Structural Econometrics of Auctions; Wage-Performance Elasticity.

JEL classification: C57 ; C13 ; D44; M52.

*We thank Mitali Das, Kartikeya Date, Xavier d'Haultfoeuille, Aureo de Paula, Paul Scott, Ali Skalli and seminar participants at CREST, the University of Paris II, the University of Luxembourg, the 2nd ECODEC Conference, and the 7th French Econometrics Conference for helpful comments and suggestions. We thank Arun Kumar for excellent research assistance. An earlier version of the paper circulated under the title "Remuneration and Performance: Evidence from Randomly Ordered Wage Auctions." Financial support from the LABEX ECODEC (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) is gratefully acknowledged.

[†]Paris School of Economics (Ecole des Ponts, ParisTech). Email: laurent.lamy78@gmail.com.

[‡]CREST-ENSAE. Email: manasa.patnam@ensae.fr.

[§]CREST-LEI and CRED, University of Paris 2. Email: michael.visser@ensae.fr.

1 Introduction

The allocation and pricing of a large range of goods and services are determined, in many instances, through an auction mechanism. Depending on the nature of the goods being auctioned, the motivation of bidders to participate may differ. In auctions of say fine wines or antiques, most bidders tend to be private collectors who wish to acquire the commodities for their own personal use; in real estate auctions, treasury auctions, or auctions of primary products such as oil and gas, most bidders are, on the contrary, banks or firms whose objective is primarily to resell the items, or to use them as intermediate products in some production process. In the latter type of auctions, it is not uncommon to observe a post-auction outcome for the winning bidders. For example, among firms that manage to purchase primary products, we may observe their subsequent production levels. Similarly, for real estate investors acquiring houses through auctions, it may be possible to record the length of time it takes before these houses are resold to their clients. Finally, in labor market auctions, an example that serves as the empirical illustration of our methodology, workers are sold to employers at salaries corresponding to the winning bids, and the post-auction outcome may be the performance or productivity of the worker.¹ In each of these examples there is potentially a relationship between the post-auction outcome and the auction price. The objective of this paper is to develop estimation methods to analyze such relationships.

The analyst faces two main challenges in causally identifying the impact of the auction determined price on some post-auction outcome. One is that prices tend to be endogenously determined by unobserved quality. Typically this quality is positively correlated with the outcome variable, resulting in an upward bias in the estimate of the price effect. The other challenge is that in auctions the allocation of items to bidders is not determined in a random way. The outcome associated with a particular item is hence not observed for an arbitrary bidder, but only for the winning bidder where the synergies between the two parties are presumably stronger. An estimation of the outcome equation which neglects this problem is then likely to suffer from selection bias.

Our econometric strategy to control for these two sources of bias is based on a control function approach.² Our empirical procedure has two steps. First, pursuing our example of a labor

¹Labor market auctions proliferate on the Internet. In these online auctions, employers can post job ads (mostly for limited-duration tasks such as translation and writing, computer programming, administrative work) and solicit bids from interested workers. Once a project is awarded to a bidder, monitoring software allows the employer to closely observe and track the productivity of the worker. See <https://sites.google.com/site/johnjosephhorton/miscellany/online-labor-markets> for examples of online labor markets wherein employers and job-seekers are matched through auctions.

²Heckman (1979) has introduced the approach to correct for selective sampling. Dubin and McFadden (1984) consider a more general setting where multiple outcome variables are subject to a selection rule. Das et al. (2003) develop a control function method which accounts for both selection and endogeneity, as a nonparametric extension of Blundell, Duncan, and Meghir (1998). The control function method is also used to account for endogeneity of explanatory variables (see Imbens and Wooldridge (2009) for an overview).

market auction, we model the auction behavior of firms under the assumption that the workers are sold one after the other, through sequential English auctions.³ In each auction, all bidders (i.e., firms) are assumed to play a strategy consisting in remaining active so long as the bidding clock is below their willingness to bid. The latter amount is referred to as the pseudo-valuation of a bidder for a given worker. It is assumed to depend on firm-worker characteristics, a set of variables capturing the history of the auctions before this worker comes up for sale (shortly referred to as “auction variables”), and an error term which we interpret as a private signal on worker productivity. Extending a classic identification result from the structural econometrics of auctions literature (see [Athey and Haile \(2008\)](#) for a survey on identification) to a setup with bidding increments, it turns out that the parameters of our auction model are non-parametrically identified if (and only if) the bidders’ private signals are independently distributed across the different bidders. The first step of our estimation procedure consists then in estimating the parameters in the pseudo-valuation equation and the distribution of private signals. Second, we assume that the expectation of the error term appearing in the outcome equation is a weighted sum of the (unobserved) private signals. The expectation of these signals can be estimated thanks to our first step estimates. The second step consists now in adding the estimated expectations as a control function in the performance equation, and then estimating the augmented equation using standard regression methods.

Our control function includes the estimated expectations of both the private signal of the winning bidder and those of the losing bidders. From an auction theory perspective, the dependence with respect to the signals of the losing bidders reflects common value components. If only the signal of the winning bidder enters the control function then it would correspond to a private value (PV) paradigm. On the contrary, if the signal of a losing bidder matters as much as the one of the winning bidder, then it would correspond to a pure common value (CV) paradigm. An interesting by-product of our model is that the two paradigms correspond to specific restrictions on the weights associated with the expected signals, and these restrictions can easily be tested.

We emphasize that our control function corrects for both the endogeneity of wages and sample selection. [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. They consider a three-equation model in which the first equation relates the outcome of interest to a single endogenous regressor and the exogenous variables, the second is a linear projection of the endogenous regressor on the exogenous variables, and the third is the selection equation which determines whether the outcome variable and the endogenous regressor are observed or not. There are two reasons why the two methodologies are not well suited for our context. One is that both frameworks can only account for sample selection bearing on a single outcome variable, not on multiple outcomes as in our case. The other reason is

³As outlined in Section 6, our methodology can be applied to other contexts and auction formats as well.

that in our model both the endogenous regressor (the wage) and the selection rule (defining which outcome variable is observed) are determined by the same variables, i.e., the pseudo-valuations of all firms. Unfortunately the setup considered by [Wooldridge \(2002\)](#) and [Das et al. \(2003\)](#) 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.

In our application we estimate a performance equation using a new and unique data set on the salaries earned by cricket players and their game-specific performances in the Indian Premier League (IPL) competition. We hereby contribute to a recent literature based on field data which tests for the presence of reciprocity ([Akerlof 1982](#)) or fairness ([Akerlof and Yellen 1990](#)) effects.⁴ [Mas \(2006\)](#) analyzes New Jersey police performance before and after final offer arbitration of wage disputes (between municipal employers and unions representing police officers). His main finding is that the number of crimes solved is higher when arbitrators judged in favor of the union, relative to when arbitrators ruled for the employer. [Lee and Rupp \(2007\)](#) exploit the fact that wages of pilots from several major US airlines companies were drastically cut in the early 2000's. They find limited evidence of reciprocity: on-time flight performance (the proxy for pilot effort) declines but only for non-bankrupt companies (pilots of such carriers may have judged the wage reductions as unjustified and unfair), and the effect is short-lived (after the first week, delays are found to be the same as before the wage-cut).

The wages in our data set are determined through a sequence of English auctions. In these auctions the bidders are the teams participating in the IPL tournament and the 'objects' for sale, the cricket players. Each player is assigned a reservation wage, which is the price at which bidding for that player should start. Players are arranged into sets, in particular according to their reservation wage and cricketing speciality. Within each set the order in which players come up for sale is randomly chosen by lottery. The winning price at which a player is sold represents the wage the auction winner has to pay to the player for participating in the tournament.

Our dataset also records detailed statistics about individual productivity. Although cricket is a team sport, the task of each player is performed more or less independently from the actions of team mates, and player-specific performances can be measured along different dimensions. The performances are observed for all matches played during the whole competition. This allows us to investigate how wages affect performance over time, and in particular whether the impact differs over the course of the tournament.

Crucial for our empirical strategy, the auction variables have an effect in the pseudo-valuation

⁴So far the studies in this branch of the literature have been based primarily on lab or field experiments. The seminal experimental contribution is [Fehr, Kirchsteiger, and Riedl \(1993\)](#). Their results provide support for the reciprocity hypothesis: experimental firms offer wages above the market-clearing wage, and employee effort increases with wages. [Gneezy and List \(2006\)](#) examine the gift-exchange hypothesis using a field experiment. They find that employees assigned to the high-wage group work harder, but this effect disappears after three hours. [Cohn, Fehr, and Goette \(2015\)](#) match the productivity of workers receiving different "wages treatment" with a post-experiment survey about subjects fairness perceptions and obtain that effort is elastic to wages only for those subjects that have fairness concerns.

equation but do not enter the outcome equation, i.e., they satisfy an exclusion restriction. The auction variables, which includes the random order of sale of the player and also the remaining budget of each firm and summary statistics on previous purchases, act as exogenous shifters in determining the willingness to bid and thus the auction outcomes (wages earned by the players, and their allocation to the different teams). Naturally, the huge wage dispersion in our environment reflects mostly heterogeneity in players ability and popularity both being valued by teams. However, part of this dispersion is due to the observable valuation shifters, which imparts features of a natural experiment in our environment, which we exploit to circumvent the endogeneity and sample selection bias that contaminates estimates from a standard regression of performance on wage.

Our results indicate a positive and statistically significant wage effect. This effect disappears, however, once we properly account for selection and endogeneity using our control function terms derived from our first stage auction model. The coefficients on the control function terms are throughout positive and significant (for the winning bidders), suggesting a positive correlation between the error terms in the performance and pseudo-valuation equations. We find, however, some limited evidence that the wage effects vary heterogeneously by the length of the tournament. Specifically, for the 2014 IPL season, we find strong positive and significant wage effects towards the end of the tournament. This result does not hold for the 2011 IPL season, possibly due to the introduction of renewable contracts in 2014.⁵

We also examine whether fairness considerations are in play for players exerting higher effort in response to higher wages. We compare whether the wage effects differ for players paid above and below their 'reference wage', defined as the average wage in some reference group. We try various possible reference groups (all players in the player's team, players with the same speciality, players having the same reservation value) and obtain no evidence in favor of the fairness channel.

The remainder of the paper is organized as follows. Section 2 presents the control function approach that enables to deal jointly with the endogeneity of the wage in the performance equation and with sample selection. Section 3 describes the crickets auctions for the Indian Premier League (IPL) and the data. Section 4 presents our empirical model, taking into account some of the specific features of the auctions as they have been organized by the IPL. Section 5 gives estimation from both stages, namely the bidding model and the performance equation and presents some robustness checks. Section 6 contains concluding remarks and discusses other settings where our methodology can be applied.

⁵With renewable contracts teams were allowed to terminate the employment of under-performing players. It could be the case that these unemployment or wage loss concerns could be more binding for players with a high wage, since they stand to lose more (compared to low wage players), which can explain why the interaction term between the week of the tournament and wage has a positive and significant effect on player performance/effort in 2014 compared to 2011 where we find no such effect.

2 The econometric strategy

This section describes our econometric strategy to control for biases due to selection and endogeneity in the performance equation. To the best of our knowledge, the method proposed here has not been developed before. It can be viewed as a combination of the insights drawn from the literature on structural econometrics of auctions and from the literature that accounts for sample selection and endogeneity using a control function approach. We deliberately present the method in a general way and abstract from some of the specific institutional details of the cricket environment. These specifics will be dealt with in Section 4.

Section 2.1 introduces the performance equation, and explains why OLS estimation can potentially lead to biased estimators. Section 2.2 describes the English auction environment with bid increments, Section 2.3 our assumptions about how firms bid in the various auctions depending on the bidding history, and Section 2.4 the control functions that correct for the various biases. Section 2.5 delineates the theoretical foundations of our approach, namely the conditions under which the bidding model and hence the control functions are non-parametrically identified. Finally, Section 2.6 develops an estimation strategy through a likelihood function approach.

2.1 The performance equation

Suppose there is a collection of workers indexed by $i = 1, \dots, N$, and a collection of firms indexed by $f = 1, \dots, F$. Throughout it is assumed that N goes large while F is fixed. Each worker-firm pair (i, f) is described by $x_{i,f}$, a vector of covariates that includes characteristics of both the worker and the firm. Let w_i be the wage earned by i and $y_{i,f}$ the performance (or productivity) of this worker, were he employed by firm f . As detailed below, w_i and the allocation of workers to firms are the outcome of a competitive bidding process. We are interested in the influence of the worker-firm variables and in particular the wage on performance and postulate the following relationship:

$$y_{i,f} = \beta_f + h(w_i) + \beta_x \cdot x_{i,f} + u_{i,f} \quad (1)$$

where β_f is a firm-specific fixed effect, h is a parametric function of the wage, β_x is a vector of parameters measuring the effects of the corresponding elements in x , and $u_{i,f}$ is an error term capturing the combined effect of all unobserved productivity determinants. The unobserved determinants may reflect variables specific to worker i and/or variables measuring synergies between worker i and firm f . In our empirical application different specifications of the function h will be considered.

To estimate β_f , β_x , and the parameters appearing in the function h , we face two kinds of

problems. The first is that estimation of the parameters is based on a selected sample. The selection arises because the wage and productivity of a worker are only observed in the data if the wage is above a player-specific threshold. This threshold is the reserve price associated to each worker and corresponds to the minimal amount under which he cannot be sold. The reserve price for i is denoted W_i^r . Therefore, if there is no bidding above W_i^r , the wage w_i and the performance of i are not recorded. In addition, if there is bidding above the reserve price, the productivity of i is only observed in the firm actually employing this worker, i.e., the firm that wins the auction for employing worker i . Letting f_i^w denote the identity of the firm to which i is matched (the index w indicating that this firm is the auction winner), only recorded is the productivity y_{i,f_i^w} ($y_{i,f}$ is unknown and counterfactual for all $f \neq f_i^w$). The parameters are therefore estimated using a sample restricted to workers whose wages are above the reserve price and who are employed by the auction winners.

The mechanism that determines whether an observation is selected in this sample is, however, not random. The matching between workers and firms results from a highly selective process: i being paired to f_i^w reveals that this firm is willing to pay more than the reserve price W_i^r and also values i more than his competitors. The likelihood of observing worker i in the estimation sample and matched with team f_i^w thus potentially increases with u_{i,f_i^w} .

Estimation of the performance equation (1) by OLS using the restricted sample leads to biased estimates because of this link between the error term and the selection rule. More precisely, there is a bias in the OLS estimates because the mean of the error term conditional on being in the sample (and given the wage and the vector of covariates), $E[u_{i,f} | w_i \geq W_i^r, f = f_i^w, w_i, x_{i,f}]$, is non-zero and varies across the observations.

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). Since the error terms $u_{i,f}$ are expected to be related positively across f (because $u_{i,f}$ captures, in part, the effect of the unobserved characteristics of worker i , and this component is presumably identical for each f), a large value of u_{i,f_i^w} most likely indicates a high degree of competition at the auction for worker i , which in turn should lead to a higher final price w_i . There is therefore potentially a dependence between the error term and the wage in the selected sample, so we suspect that $E[u_{i,f} | w_i \geq W_i^r, f = f_i^w, w_i, x_{i,f}]$ increases with w_i .

If the endogeneity of wages were the only source of bias (i.e., in the absence of sample selection),⁶ a standard instrumental variable method could be used to solve the problem provided a suitable instrument is available. Whenever it is necessary to account jointly for endogeneity and sample selection (as in our case and a fortiori with a selection process of a multidimensional nature), a pure IV approach seems unfeasible.

⁶Sample selection does not play a role in our model if there are no reserve prices and if all workers are somehow randomly assigned to firms (which is the case if, for example, each worker is valued equally by the different firms independently of the set of workers already hired).

2.2 The auction environment

The N workers are auctioned one after the other, through sequential English auctions with a publicly observed reserve price, still denoted as W_i^r . Unlike earlier econometric auction papers, we do not assume that the data are generated by the button English auction where the bidding clock rises continuously.⁷ Instead, we explicitly account for bid increments in the auction environment. We feel that it is important to do so because bid increments are the norm in real-world implementations of English auctions. In most auction houses throughout the world, for instance, goods are sold via oral ascending auctions and discontinuities are observed in the bidding clocks because auctioneers invite bidders, at each given current price, to submit new bids equal to the current price plus a positive increment. Furthermore, the magnitude of bid increments is in practice typically far from negligible. In our application on cricket auctions, for example, they represent between 5 and 10% of the current price.⁸

An increment rule is characterized by a function $\Delta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, with $\inf_{x \in \mathbb{R}_+} \Delta(x) > 0$. For the moment no additional assumptions need to be made on the function Δ . Given the increment rule, the English auction is assumed to operate as follows: Bidding for worker i starts at $W_i^r \geq 0$. If no firm enters the auction when the bidding clock is at this value then i remains unsold. If a single firm enters, then i is sold at the reserve price. If at least two bidders have entered, they are asked simultaneously whether they wish to submit a new bid at the price $W_i^r + \Delta(W_i^r)$. Three possibilities must now be distinguished. First, if none of the active bidders overbids, then the tie-breaking rule consists in picking randomly and with equal probability the winner among the active bidders at W_i^r , and the winning firm has to pay W_i^r to acquire the worker. Second, if a single bidder overbids then this bidder wins the auction and has to pay W_i^r . Third, if at least two bidders overbid the clock rises to $W_i^r + \Delta(W_i^r)$, and the process just described repeats itself. In this last case, only the firms which have submitted bids at $W_i^r + \Delta(W_i^r)$ are allowed to proceed with the auction, i.e., we assume that exit from the auction is irreversible. To summarize, we are considering an English auction environment with irrevocable exits and a discontinuously rising clock. If, at some point, the clock has reached the price x , the bidding process either terminates at this value (when at most one firm is willing to pay $x + \Delta(x)$), or the process continues and the clock jumps to $x + \Delta(x)$ (when at least two firms are willing to pay this amount to purchase the worker).

No assumptions need to be made on the auction paradigm underlying the bidding behavior. It is thus not required to assume that bids are generated for instance by a private value (PV) model, or by a common value (CV) model. Given that we are dealing with English auctions with

⁷Examples of papers adopting the button English auction framework are [Baldwin et al. \(1997\)](#), [Paarsch \(1997\)](#), and [Athey and Haile \(2002\)](#).

⁸In a similar vein, [Hickman \(2010\)](#) points out that increments should not be neglected in electronic ascending proxy auctions as on eBay. Although bid increments are small in his data set (around 1 and 2%), the probability that the difference between the highest and the second-highest proxy bids (the analog of the two highest pseudo-valuations in our setup) is less than the bid increment ranges between 20 and 25%.

irrevocable exit, the strategy of firm f consists in remaining active in the auction for worker i so long as the bidding clock is below $V_{i,f}$, the maximal amount f is prepared to bid to acquire i .⁹ We will refer to this amount as the pseudo-valuation of firm f for this worker.¹⁰ It is allowed to depend on the information released to all bidders during the auctions prior to the sale of worker i .

Note that since the increments are bounded away from zero, the auction necessarily ends in a finite number of iterations (in particular because the auction price will never go beyond the second-highest pseudo-valuation). Note also that when the increments tend to zero, the final outcome “converges” to the outcome obtained with the button English auction where the winner is the bidder with the highest pseudo-valuation and the final price coincides with the maximum of the second-highest pseudo-valuation and the reserve price.

Crucially, the auction data base is assumed to record the reserve price for all workers, regardless of whether they are sold or not. If worker i is sold we observe in addition the auction winner f_i^w and the corresponding wage w_i , but otherwise these variables are not defined. This is typically the (minimal) kind of information available in most auction data sets. What we learn from the auction data is thus the following: If worker i is not sold we know that all firms have a pseudo-valuation below the reserve price: $V_{i,f} < W_i^r$ for all f . If i is instead sold, then the winning price is either equal to the reserve price, $w_i = W_i^r$, or strictly above it, $w_i > W_i^r$. The first case occurs when at least one firm has entered the auction, and at most one firm remains active at the price $W_i^r + \Delta(W_i^r)$. This is equivalent to $V_{i,f_i^w} \geq W_i^r$ and $V_{i,f} < W_i^r + \Delta(W_i^r)$ for all $f \neq f_i^w$. The second case occurs when there are at least two firms actively bidding up to the winning price w_i , and at most one firm remains active at the price $w_i + \Delta(w_i)$. We then know that the pseudo-valuation of the winner exceeds the winning price, all other firms have pseudo-valuations below the winning price plus the increment, and the pseudo-valuation of at least one other firm (a bidder active at w_i) is between the winning price and the winning price plus the increment. Formally, this is equivalent to $V_{i,f_i^w} \geq w_i$, $V_{i,f} < w_i + \Delta(w_i)$ for all $f \neq f_i^w$, and $V_{i,f} \in [w_i, w_i + \Delta(w_i))$ for at least one $f \neq f_i^w$.¹¹ Note that our formulation of the two cases accounts for the possibility of ties (at respectively $w_i = W_i^r$ and $w_i > W_i^r$).

⁹This is true if bidders do not observe the number of competing active bidders and their identities. Otherwise, our model makes the restriction that bidders do not take this information into account in their bidding behavior.

¹⁰We call $V_{i,f}$ the ‘pseudo-valuation’ instead of ‘valuation’ since the latter expression may wrongly give the impression that firms bid as if each worker is sold in an isolated, stand-alone auction. In sequential auctions, $V_{i,f}$ reflects that other workers are sold after i : For both strategic reasons and the presence of either substitutabilities or complementarities between workers, the pseudo-valuation may differ from what the firm would have bid had i been the unique worker to be sold.

¹¹Without further information, we cannot tell exactly for which $f \neq f_i^w$ we have $V_{i,f} \in [w_i, w_i + \Delta(w_i))$.

2.3 The (pseudo-)valuation model

Next we turn to the specification of the pseudo-valuations $V_{i,f}$. We assume that $V_{i,f}$ depends not only on the firm-worker characteristics $x_{i,f}$, but also on a vector of covariates $z_{i,f}$, the auction variables. This vector captures all relevant variables observed by f just before i is being auctioned. It contains in particular the order of sale of i in the auction sequence (whether this worker comes up for sale first, or second, etc.), the remaining budget of firm f at this stage, and information on the characteristics of the previous workers purchased by this firm. By including the information on previous purchases, we have in mind that if a firm has already bought several workers of a given type, then buying yet an additional one is less valuable (because of substitutability between workers of a given type). Then we assume that $V_{i,f}$ has the following form

$$\log[V_{i,f}] = G_f(z_{i,f}, x_{i,f}) + \epsilon_{i,f} \quad (2)$$

where $\epsilon_{i,f}$ is an error term capturing the unobserved valuation-determinants, and G_f , $f = 1, \dots, F$, are unspecified functions of the explanatory variables. We will interpret $\epsilon_{i,f}$ as the private signal received by f on the worker productivity of i . It captures partly how f evaluates the ability of worker i (i.e., a pure common value feature shared by all firms), and also how this firm anticipates the worker will perform and produce in its specific workplace environment (i.e., an idiosyncratic synergy between i and f). For our identification result, we assume that the vector $(\epsilon_{i1}, \dots, \epsilon_{iF})$ is distributed according to a continuously differentiable density which has support \mathbb{R}^F .

In the empirical application parametric functions are chosen for each G_f . However, to formulate our identification result and state the conditions under which it holds, it is convenient to adopt a nonparametric setting in which the functions are left unspecified. The distinction between the variables $x_{i,f}$ and $z_{i,f}$ does not play any role for identification but will be crucial from an empirical perspective (to gain power in estimating the control functions): the former variables do enter the performance equation while the latter ones satisfy an exclusion restriction. We are thus assuming that the characteristics of firms and workers matter for performance, but not the variables such as the order of sale or the remaining budget of the firm.

2.4 Control function

In section 2.1 we explained that by estimating the regression model

$$y_{i,f_i^w} = \beta_{f_i^w} + h(w_i) + \beta_x \cdot x_{i,f_i^w} + u_{i,f_i^w}, \quad (3)$$

using data from the selected sample of workers actually sold, biased estimates are obtained because $E[u_{i,f_i^w} | w_i \geq W_i^r, f = f_i^w, w_i, x_{i,f_i^w}] \neq 0$. The control function approach consists in modeling $E(u_{i,f_i^w} | \mathcal{I})$ where \mathcal{I} denotes the full information set, i.e., it contains all variables ob-

served ex post by the econometrician from all N auctions. The set \mathcal{S} thus includes in particular $z_{i,f}$, $x_{i,f}$, and W_i^r for all i and f , and w_i and f_i^w for all workers i actually sold. The resulting expectation is then added to (3) to control for the sample selection and the endogeneity of w_i . We will model the control function $E(u_{i,f_i^w}|\mathcal{S})$ by using our auction model. The control function is therefore not specified in some ad hoc way but explicitly draws on auction theory. In this respect our approach is reminiscent of Dubin and McFadden (1984) who construct their control function by drawing on discrete choice theory.¹²

By conditioning on \mathcal{S} we condition not just on $(w_i \geq W_i^r, f = f_i^w, w_i, x_{i,f_i^w})$ but also on a set of additional variables. This is necessary because our auction model implies that the control function should depend on various additional variables, in particular the values of z and x for all firms. Adding other variables is also of practical empirical importance: it avoids the problem of having a high collinearity between the regressors w_i and x_{i,f_i^w} appearing in (3), and the variables in the control function.

Letting $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iF})$, we can write (by iterated expectations) $E(u_{i,f_i^w}|\mathcal{S}) = E_{\epsilon_i}(E(u_{i,f_i^w}|\mathcal{S}, \epsilon_i)|\mathcal{S})$, where the first expectation (after the equality sign) is with respect to ϵ_i given \mathcal{S} , and the second is with respect to u_{i,f_i^w} given \mathcal{S} and ϵ_i . The second expectation is assumed to satisfy the following key assumption.

$$\mathbf{A1:} \quad E(u_{i,f_i^w}|\mathcal{S}, \epsilon_i) = \gamma \cdot \epsilon_{i,f_i^w} + \chi \cdot \sum_{f \neq f_i^w} \epsilon_{i,f}.$$

Assumption A1 states that given \mathcal{S} and ϵ_i , the expectation of u_{i,f_i^w} only depends on ϵ_i . This kind of assumption is typically made in the control function literature (see Imbens and Wooldridge (2009)).¹³ Under A1 the signals received by all F firms are correlated with u_{i,f_i^w} , and hence with the subsequent performance of i in firm f_i^w . The signals of all losing bidders are assumed to affect unobserved worker productivity in the same way (the coefficient associated with each $\epsilon_{i,f}$ is χ). This symmetry assumption is not necessary for identification purposes but is made for practical reasons (it is difficult to estimate with precision a separate coefficient for each losing bidder). We then interpret $\gamma > \chi = 0$ as reflecting that the underlying auction is a pure PV auction: conditional on \mathcal{S} , signals reflect solely idiosyncratic synergies between firms and workers. On the contrary, $\gamma = \chi > 0$ indicates that the underlying auction has the flavor of a pure CV auction in which each (symmetric) bidder receives independently a private signal.

¹²Interestingly, Dubin and McFadden's (1984) discrete choice model and our auction model are closely related. The two models and resulting control functions have the same structure when $h(w_i) = W_i^r = 0$ for all i , if there are no increments (i.e., the bidding clock rises continuously), and when wages are unobserved in the data. In this case there is no longer an endogenous variable in (3), and the sample selection rule is defined like in Dubin and McFadden: y_{i,f_i^w} is observed if $V_{i,f_i^w} = \max_f V_{i,f}$. Compared to our setup, Dubin and McFadden put additional structure on the error terms $u_{i,f}$ and $\epsilon_{i,f}$. They assume that $\epsilon_{i,f} = \epsilon_i + \epsilon_f$, and $u_{i,f} = \epsilon_i$, and hence $\epsilon_{i,f}$ and $\epsilon_{i,f'}$ are dependent variables in their case. As shown below, however, the parameters of the auction model are only identified non-parametrically under independence of these error terms across firms.

¹³A1 differs from what is typically assumed in this literature in the sense that the error term u_{i,f_i^w} is allowed to be correlated not just with ϵ_{i,f_i^w} (correlation between errors of the same unit of observation) but also with $\epsilon_{i,f}$ for $f \neq f_i^w$ (correlation between error terms of different units) if $\chi \neq 0$.

In general, we expect that the weight of the signal of the winning bidder in the performance equation should be larger than the one of the losing bidders, i.e., $\gamma \geq \chi \geq 0$. Rejecting from bidding data alone the private value paradigm (against common values) is known to be a difficult problem without parametric restrictions.¹⁴ There is a key ingredient here that allows us to circumvent this problem: similarly to [Hendricks et al. \(2003\)](#), we observe an ex-post performance measure of each worker.

Under A1, modeling the control function $E(u_{i,f}^w | \mathcal{D})$ amounts to propose a model for the expectation $E(\epsilon_{i,f} | \mathcal{D})$, for all f . This last expectation depends in particular on the joint distribution of the signals $\epsilon_{i,1}, \dots, \epsilon_{i,F}$. The next section establishes the conditions under which this distribution is identified from the bidding data.

2.5 Identification

The primitives of our auction model are the functions G_f , $f = 1, \dots, F$, on the support of the covariates (note that we do not exclude that covariates take discrete values) and the joint distribution of the vector of signals $(\epsilon_{i,1}, \dots, \epsilon_{i,F})$ on \mathbb{R}^F . Although we develop in [Section 2.6](#) a fully parametric estimation approach, the aim below is first to discuss under which restrictions our auction model is non-parametrically identified. The following proposition is the analog of [Theorem 6.2.1](#) in [Athey and Haile \(2008\)](#), but in a framework with reserve prices and, more importantly, bidding increments. To extend their result, we require the reserve prices to vary on the interval $[0, \Delta(0)]$ in order to guarantee that the distribution of wages has full support on \mathbb{R}_+ . We also require the increment rule Δ to satisfy an additional condition, namely that $w \rightarrow w + \Delta(w)$ is strictly increasing and continuous in w . If Δ satisfies this condition we say that it is regular.

Proposition 1: Suppose that the increment rule is regular. Suppose also that the number of bidders that is eligible to bid varies exogenously across the N auctions, and that there exists a vector of covariates $(z^*, x^*) = ((z_f^*, x_f^*))_{f=1, \dots, F}$ such that the support of W^r conditional on (z^*, x^*) contains the interval $[0, \Delta(0)]$.

Then the auction model is non-parametrically identified from the data if i) the vectors of signals $(\epsilon_{i,1}, \dots, \epsilon_{i,F})$ are i.i.d. across i ; ii) $\epsilon_{i,f}$ is independent from $\epsilon_{i,f'}$ for each i and $f \neq f'$, and $\epsilon_{i,f}$ is distributed according to the distribution H_f with $E[\epsilon_{i,f}] = 0$;¹⁵ iii) $\epsilon_{i,f}$ is independent from (z_i, x_i, W_i^r) .

The proof is in [Appendix A.1](#).

We postpone our comments on the proposition until the end of this subsection, and turn now to the calculation of $E(\epsilon_{i,f} | \mathcal{D})$. Under the three identifying assumptions listed in the proposition

¹⁴[Laffont and Vuong \(1996\)](#) establishes a non-identification result for standard auctions.

¹⁵The condition $E[\epsilon_{i,f}] = 0$ is a location normalization.

(our working hypotheses from now on), it turns out that this expectation can be expressed in a relatively simple way. To show this, we first introduce some additional notation. Let $G_{i,f} = G_f(z_{i,f}, x_{i,f})$, $\bar{H}_{i,f} = H_f(\log(w_i) - G_{i,f})$ and $\bar{H}_{i,f}^\Delta = H_f(\log(w_i + \Delta(w_i)) - G_{i,f})$; Furthermore, let S represent the set of active bidders at the winning price w_i , and $|S|$ the cardinality of this set.

Given that f_i^w is the winner, we must have $S \supseteq f_i^w$. Furthermore, if $w_i > W_i^r$, we must also have $|S| \geq 2$ (since if $|S| = 1$ the winning bid should necessarily have been lower); If instead $w_i = W_i^r$ then $|S| \geq 1$ (the auction winner is possibly the only active bidder at the reserve price). We now have to distinguish two cases: $|S| \geq 2$ and $|S| = 1$. When $|S| = 1$, we necessarily have $w_i = W_i^r$ and by definition f_i^w is the only active bidder at this price. We then know that f_i^w has a pseudo-valuation above w_i , and all other bidders have a pseudo-valuation below w_i . When $|S| \geq 2$, and regardless of whether the winning bid is strictly above or equal to the reserve price, two events may have occurred: either firm f_i^w has a pseudo-valuation above $w_i + \Delta(w_i)$ (and has won by overbidding at price $w_i + \Delta(w_i)$ if $|S| \geq 2$), or f_i^w has a pseudo-valuation in the interval $[w_i, w_i + \Delta(w_i))$ (and has won by winning the tie at price w_i if $|S| \geq 2$). The probabilities of these two events are respectively denoted $p_i^{over}(\mathcal{S}, S)$ and $p_i^{tie}(\mathcal{S}, S)$. From Bayesian updating, we obtain that

$$p_i^{over}(\mathcal{S}, S) = \frac{1 - \bar{H}_{i,f_i^w}^\Delta}{(1 - \bar{H}_{i,f_i^w}^\Delta) + \frac{1}{|S|}(\bar{H}_{i,f_i^w}^\Delta - \bar{H}_{i,f_i^w})} \quad (4)$$

and $p_i^{tie}(\mathcal{S}, S) = 1 - p_i^{over}(\mathcal{S}, S)$.¹⁶

To proceed, let $E[\epsilon_{i,f} | \mathcal{S}, S]$ denote the expectation of $\epsilon_{i,f}$ conditional on \mathcal{S} and the fact that the set of active bidders at the winning price w_i is $S \supseteq f_i^w$. Given the identifying assumptions conditions on the error terms ϵ , our auction environment (Section 2.2), and the specification of the pseudo-valuations (Section 2.3), this expectation can be written as

$$E[\epsilon_{i,f} | \mathcal{S}, S] = E[\epsilon_{i,f} | \epsilon_{i,f} < \log(w_i) - G_{i,f}] \text{ if } f \notin S, \quad (5)$$

and

$$E[\epsilon_{i,f} | \mathcal{S}, S] = E[\epsilon_{i,f} | \epsilon_{i,f} \in [\log(w_i) - G_{i,f}, \log(w_i + \Delta(w_i)) - G_{i,f}]] \text{ if } f \in S \setminus \{f_i^w\}. \quad (6)$$

These expressions are valid for any subset S . For the auction winner, we separately express this expectation depending on whether f_i^w is the only bidder in S or not. For $|S| \geq 2$ we have

¹⁶The denominator corresponds to the probability that the firm f_i^w wins at price w_i given that the firms in the set $S \setminus \{f_i^w\}$ are prepared to bid up to w_i , but not beyond $w_i + \Delta(w_i)$.

$$\begin{aligned}
E[\epsilon_{i,f_i^w} | \mathcal{S}, S] &= p_i^{over}(\mathcal{S}, S) \cdot E[\epsilon_{i,f_i^w} | \epsilon_{i,f_i^w} \geq \log(w_i + \Delta(w_i)) - G_{i,f_i^w}] \\
&\quad + p_i^{tie}(\mathcal{S}, S) \cdot E[\epsilon_{i,f_i^w} | \epsilon_{i,f_i^w} \in [\log(w_i) - G_{i,f_i^w}, \log(w_i + \Delta(w_i)) - G_{i,f_i^w}]],
\end{aligned} \tag{7}$$

while for $|S| = 1$ the expectation simplifies to (using that in this case $w_i = W_i^r$)

$$E[\epsilon_{i,f_i^w} | \mathcal{S}, S] = E[\epsilon_{i,f_i^w} | \epsilon_{i,f_i^w} \geq \log(W_i^r) - G_{i,f_i^w}]. \tag{8}$$

For any $S \supseteq f_i^w$, we now let $p_i^S(\mathcal{S})$ denote the probability that the set of bidders who are active at price w_i is S , conditional on \mathcal{S} . In Appendix A.2 we give the precise expression of this probability.

All elements to compute the expectation $E[\epsilon_{i,f} | \mathcal{S}]$ are now available. Indeed, by iterated expectations, we obtain

$$E[\epsilon_{i,f} | \mathcal{S}] = \sum_{\substack{S \subseteq \{1, \dots, F\} \\ f_i^w \in S}} p_i^S(\mathcal{S}) \cdot E[\epsilon_{i,f} | \mathcal{S}, S]. \tag{9}$$

We see that $E[\epsilon_{i,f} | \mathcal{S}]$ can be expressed as a weighted sum of the expectations $E[\epsilon_{i,f} | \mathcal{S}, S]$ for the different sets S . The latter expectations and the weights $p_i^S(\mathcal{S})$ only depend on the functions G_f and H_f , for all f . Since these functions are identified under the three conditions of the proposition, $E[\epsilon_{i,f} | \mathcal{S}, S]$ and $p_i^S(\mathcal{S})$ are also identified. Hence we can identify the expectation $E[\epsilon_{i,f} | \mathcal{S}]$ for each f , which in turn implies that we can identify the control function $E(u_{i,f_i^w} | \mathcal{S})$.

Before ending this subsection, we wish to make several comments on our proposition. 1) An important message of the proposition is that the auction model is identified non-parametrically under independence of the unobserved signals. On the contrary, identification is not achieved if we allow for any form of correlation between $\epsilon_{i,f}$ and $\epsilon_{i,f'}$ for any $f \neq f'$. This is a theoretical justification for imposing the identifying assumptions in our empirical application. 2) Crucial for the identification result to hold is that the wage is observed for sold workers. In the absence of this information identification is not achieved. This means in particular that the model of Dubin and McFadden (1984) is not non-parametrically identified, and identification is only obtained in their case through parametric restrictions. 3) Proposition is in the same vein as Theorem 6.2.1 of Athey and Haile (2008). There are two difference with their setting: first they do not consider binding reserve prices, and, more importantly, they do not consider bid increments. A priori, increments preclude pointwise identification. To get identification, it is necessary to assume that the support of W^r conditional on (z^*, x^*) contains the interval $[0, \Delta(0)]$. This amounts to assuming that there is some exogenous variation in the reserve price.

2.6 Estimation: a likelihood approach

We develop a parametric two-step estimation method. In the first step we estimate the primitives of the auction model by maximizing the likelihood function using the auction data, i.e., $(z_{i,f}, x_{i,f}, W_i^r)$ for all i and f , and in addition (f_i^w, w_i) for workers who are sold. We parametrize the functions $(G_f, H_f)_{f=1,\dots,F}$ by some parameter $\alpha \in \mathbb{R}^L$. To emphasize that these functions depend on α , they will be denoted $(G_f^\alpha, H_f^\alpha)_{f=1,\dots,F}$. Let $\hat{\alpha}$ be the ML estimator of α , and $(G_f^{\hat{\alpha}}, H_f^{\hat{\alpha}})_{f=1,\dots,F}$ the corresponding first-stage consistent estimators of the primitives. Furthermore, let $CF_{i,f}^\alpha = E[\epsilon_{i,f} | \mathcal{I}]$ and $CF_{i,f}^{\hat{\alpha}}$ the associated consistent estimator of this expectation. Finally, we obtain the estimated control function $\hat{E}(u_{i,f_i^w} | \mathcal{I}) = \gamma \cdot CF_{i,f_i^w}^{\hat{\alpha}} + \chi \cdot \sum_{f \neq f_i^w} CF_{i,f}^{\hat{\alpha}}$.

In the second step we add the estimated control function in (3), and estimate by OLS the corrected performance equation:

$$y_{i,f_i^w} = \beta_{f_i^w} + h(w_i) + \beta_x \cdot x_{i,f_i^w} + \gamma \cdot CF_{i,f_i^w}^{\hat{\alpha}} + \chi \cdot \sum_{f \neq f_i^w} CF_{i,f}^{\hat{\alpha}} + \text{error}_{i,f_i^w} \quad (10)$$

using the complete data set (auction data and performance data) from the sample of workers sold at auction. This second step allows to estimate consistently all remaining parameters, i.e., $\beta_{f_i^w}$, the parameter(s) in h , β_x , γ , and χ .

We now turn to the expression of the likelihood function. Any worker i is either unsold, sold to firm f_i^w at the reserve price W_i^r , or sold to f_i^w at the price $w_i > W_i^r$. For a given α , the probabilities of these three events, conditional on x_i, z_i , and the reserve price W_i^r , are denoted by $L_i^1(\alpha)$, $L_i^2(f_i^w; \alpha)$, $L_i^3(w_i, f_i^w; \alpha)$.

The probability that i remains unsold is the probability that all bidders have a pseudo-valuation below W_i^r . Under the independence of the ϵ s we thus have

$$L_i^1(\alpha) = \prod_{f=1}^F H_f^\alpha(\log(W_i^r) - G_f^\alpha(z_{i,f}, x_{i,f})).$$

To express the likelihood contributions of the two other types of workers, we denote $\bar{H}_{i,f}(\alpha) = H_f^\alpha(\log(w_i) - G_f^\alpha(z_{i,f}, x_{i,f}))$ and $\bar{H}_{i,f}^\Delta(\alpha) = H_f^\alpha(\log(w_i + \Delta(w_i)) - G_f^\alpha(z_{i,f}, x_{i,f}))$.

The probability to be sold at the reserve price to bidder f_i^w is the probability that all bidders not belonging in the set S (with $|S| \geq 1$) have a pseudo-valuation below W_i^r , all bidders in S except f_i^w have pseudo-valuations between W_i^r and $W_i^r + \Delta(W_i^r)$, and f_i^w either has a pseudo-valuation above $W_i^r + \Delta(W_i^r)$ or has a valuation between W_i^r and $W_i^r + \Delta(W_i^r)$ and has won the tie (which occurs with probability $\frac{1}{|S|}$):

$$L_i^2(w_i, f_i^w; \alpha) = \sum_{\substack{S \subseteq \{1,\dots,F\} \\ f_i^w \in S, |S| \geq 1}} \prod_{f \notin S} \bar{H}_{i,f}(\alpha) \cdot \prod_{f \in S \setminus \{f_i^w\}} (\bar{H}_{i,f}^\Delta(\alpha) - \bar{H}_{i,f}(\alpha)) \cdot \left((1 - \bar{H}_{i,f_i^w}^\Delta(\alpha)) + \frac{1}{|S|} \cdot (\bar{H}_{i,f_i^w}^\Delta(\alpha) - \bar{H}_{i,f_i^w}(\alpha)) \right).$$

Since the composition of the set S is unknown, we sum in the above expression over all possible sets S which include f_i^w and with $|S| \geq 1$.

Finally we consider the contribution to the likelihood of a worker sold above the reserve price. The probability L_i^3 that i is sold to bidder f_i^w at $w_i > W_i^r$ is analogous to the contribution L_i^2 except that now only sets with at least two bidders should be considered, i.e., the constraint $|S| \geq 1$ in the previous summation should be replaced by $|S| \geq 2$.

Letting K_1 be the set of workers unsold, K_2 the set of workers sold at the reserve price, and K_3 the set of those sold above the reserve price, the likelihood function can be written as

$$L(\alpha) = \prod_{i \in K_1} L_i^1(\alpha) \times \prod_{i \in K_2} L_i^2(f_i^w; \alpha) \times \prod_{i \in K_3} L_i^3(w_i, f_i^w; \alpha). \quad (11)$$

The maximum likelihood estimator $\hat{\alpha}$ is the value of α that maximizes $L(\alpha)$.

As far as we know, our paper is the first to develop a likelihood function based on the English auction with bid increments. In the previous literature (see e.g. [Baldwin et al. \(1997\)](#), [Paarsch \(1997\)](#) and [Paarsch and Hong \(2006\)](#)), statistical inference has been based on likelihood functions associated with the button English auction. The button English auction version of our likelihood function is obtained by letting the increments in the above expressions go to zero. For instance, $L_i^3(w_i, f_i^w; \alpha)$ converges to (up to the multiplication constant $\log(1 + \frac{\Delta(w_i)}{w_i})$, which does not depend on α):

$$\sum_{\substack{f=1 \\ f \neq f_i^w}}^F \prod_{\substack{f'=1 \\ f' \neq f, f_i^w}}^F \bar{H}_{i,f'}(\alpha) \cdot \bar{h}_{i,f}(\alpha) \times (1 - \bar{H}_{i,f_i^w}(\alpha)), \quad (12)$$

where $\bar{h}_{i,f}(\alpha) := (H_f^\alpha)'(\log(w_i) - G_f^\alpha(z_{i,f}, x_{i,f}))$, the derivative of H_f^α evaluated at $\log(w_i) - G_f^\alpha(z_{i,f}, x_{i,f})$. The likelihood functions appearing in the aforementioned papers are based on analogues of expression (12): it corresponds to the density associated with the event where all bidders f' ($f' \neq f, f_i^w$) have pseudo-valuations below w_i , bidder f has a pseudo-valuation exactly equal to w_i , and the winner has a pseudo-valuation exceeding w_i . Since the identity of the bidder f is unknown, the summation is over all possible f .

3 Data & Environment

3.1 Tournament and player performance

The Indian Premier League is an annual cricket tournament organized by the Board of Cricket Control in India (BCCI). The tournament involves eight to ten teams that compete by playing

matches in a double round-robin format lasting six weeks. At the end of this first stage, the four best ranked teams compete during one week in a playoff to determine the final winner of the tournament. In the tournament, a cricket match is played over a set time period¹⁷ between two teams consisting of 11 players (of mixed nationalities) who are selected from the team squads. The size of the squads is not fixed and consists of about two dozen of players.¹⁸ In our empirical analysis we focus on two seasons of the IPL tournament, played in 2011 and 2014, for two specific reasons. First, 2011 and 2014 represent years in which major player auctions were held, whereby all players were (re)allocated to teams. Second, there was a significant shift in the structure of player contracts between 2011 and 2014, from fixed-term to renewable. We are interested in exploiting this contractual variation to examine how the wage effect (if any) varies with incentives.

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 set of individual measures of performance that are idiosyncratic and largely independent of how other team members perform. Players are categorized into four categories: batsman, bowler, wicket keeper and all-rounder.¹⁹ The general composition of a cricket squad 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). In our data, among sold players, the proportions are 26% batsmen, 41% bowlers, 12% wicket-keepers and 22% all-rounders.

We observe a wide array of performance measures for a cricket player depending on his speciality. In our data we obtain, for each match in the tournament, a variety of player-specific performance indicators and combine them into a multidimensional index of performance using factor analysis. As the performance indicators tend to be highly correlated with each other, we use the method of principal components to extract a single variable that captures most of the variation amongst these indicators. We discuss in the Appendix, the construction of our aggregate performance index and provide a description of each speciality-specific performance indicator in our data.

Data sources: We obtain performance data on all matches played in the tournament from

¹⁷In the IPL, a match is generally completed in 3 hours. The match involves one team batting (striking the ball) first 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.

¹⁸The average number of players per squad were 22.9 and 26.8 in 2011 and 2014, respectively.

¹⁹A 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.

www.espncricinfo.com. All data on player auctions, described in the following section, are manually compiled from the recordings and minutes²⁰ of the (publicly-broadcast) auction proceedings.

3.2 Player auctions

Beginning in 2008, once every three years, the IPL organizes auctions to (re)allocate players to teams. These auctions offer an opportunity for teams to buy their players through a centralized market. The format of sale consists of a sequential procedure whereby players are sold sequentially through an English (or ascending) auction with a public reserve price. In each of those independent auctions, the winning bid represents the player's salary for the tournament. In 2014, the auctions are the only way to hire new players so that those remaining unsold in the auctions do not participate in the tournament. By contrast, in 2011 about half of the squad is purchased outside the auctions. The number of teams/bidders and players that participate in the auction varies across auctions years. In 2011 the auction consisted of 10 bidders and 350 players. In 2014, the auctions consisted of 8 teams/bidders and 514 players. Given that only a subset of players from the initial list of participating players were auctioned,²¹ we are left at the end with 333 and 317 player auctions with 111 and 122 players sold, for 2011 and 2014 respectively. The remaining players remain unsold as they received no bid above their reservation price. For the sold players, we observe player-specific performance measures across all matches in the tournament.

Player Contracts: In the 2014 auction, player wage contracts²² were fixed for a one-year term with the option of renewal for an additional one or two years. Players whose contracts were terminated at the end of the first year would be pooled into a mini-auction in the subsequent year and re-allocated across teams through this auction. In contrast, the 2011 tournament offered players a three-year fixed term contract. As a result, the 2014 auctions and tournament present an ideal setting for our analysis as we are able to examine player performance in the first season following the auction, where players face a genuine incentive to perform that is effective immediately after their wage-determination.²³

²⁰The minutes of the live auction proceedings were obtained from ESPN-Cricinfo (<http://www.espncricinfo.com/indian-premier-league-2014/content/story/718095.html>).

²¹More precisely, the final phase of the auction was reserved for an 'accelerated' auction whereby teams would pick their 'wish-list' of players from the un-auctioned remaining lot, that they would like to bid for. This list could also include the set of unsold players that teams would like to bring back for bidding. In our analysis, we drop players that appeared in the accelerated auction phase, as their appearance depends endogenously on each bidder's preferences that our model does not account for. In 2011, all 28 unsold players were brought back to be auctioned, out of which 12 players were sold. In 2014, 244 players remained unsold; at the request of the teams, 113 unsold players were brought back to be auctioned, out of which 29 players were finally sold.

²²All player salaries are taxed in India; however overseas players face a (uniformly) lower tax burden on their salary compared to Indian players (approximately 10% compared to 40%). In the analysis, our measure of wages are in gross terms, but we account for the tax-differential by including a dummy for Indian players.

²³For 2011 auction, it is likely that the fixed-term structure of the wage contract dampened player incentives to

3.2.1 Auction Regulations

Prior to the auction, 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.²⁴

Teams faced a set of explicit rules with regard to both team composition and bidding behavior. Table 1 presents an overall summary of all the auction features both in 2011 and 2014. These rules play an important role in determining some constraints that bidders face whilst bidding. Some important bidder constraints, based on the rules, are as follows:

Spending cap: In order to encourage a balanced competition, the organizers effectively impose a spending cap on the total amount that any bidder is allowed to spend in the auction. The spending cap allocated to a bidder depends on the number of players retained by the team from its previous year’s squad. Teams are allowed to retain a maximum of four (2011) or five (2014) players from their previous year’s squad, and the spending cap decreases with the number of retained players (see Table 1).

Overseas player quota: The organizers impose a limit on the number of overseas (non-Indian) players in any team. For example, in 2014, a team could only purchase a maximum number of 9 overseas players. As a result of this constraint, bidders tend to find Indian players more valuable *ceteris paribus*. In 2014, 44% (resp. 66%) of the auctioned (resp. sold) players are Indian, whereas in 2011, 14% (resp. 39%) of the auctioned (resp. sold) players are Indian.

Right to match options: In 2014, the organizers introduced a special feature, whereby they allocated a limited number of so-called ‘Right to Match (RTM)’ options to teams. A RTM card allows a team to buy back a player from their previous year’s squad by matching the player’s winning bid when he is sold at the auction. The number of allocated cards depends on the number of retained players.²⁵

Note that these constraints evolve dynamically depending on the auction sequence and the teams’ cumulative purchases up till that point. For instance, the budget constraint just before a given player is auctioned (initial spending cap minus the amount of money spent by the team before this player came up) becomes more binding as the auction proceeds and teams purchase their share of players. The speciality and nationality constraints may also start to affect each bidder’s valuations depending on the speciality and nationality of the already enlisted players. In the next section we discuss how to incorporate these constraints using variables that capture

perform until the very last season, where they eventually faced the threat of re-allocation. While it is possible to analyse performance outcomes from the last season, we miss the immediate effect of the sudden change in wage experienced by players in the first season following the respective auctions.

²⁴In 2014, seven different reserve prices were used (1,2,3,5,10,15 and 20 Millions of Rupees). In 2011, six different reserve prices were used (20,50,100,200,300 and 400 thousands of dollars).

²⁵An RTM card is equivalent to what is called the “right-of-first-refusal” option in the auction literature (see Bikhchandani, Lippman, and Ryan (2005)).

the cumulative purchases and budget of each team at each point of the auction procedure.

Table 2 presents a summary of bidder purchases in the 2014 and 2011 auctions. Prior to the auctions, teams retained, on average, 3 players in 2014 and 1.2 players in 2011. These players (a total of 24 and 12 retained players in 2014 and 2011 respectively) do not appear in the auctions. The average spending cap for teams was around 5 million USD in 2014 and 7 million in 2011. Most teams exhaust almost all of their budget at the end of the auction (for instance, the average amount of money left unspent at the end of the 2014 sale was 0.5 million USD). Finally, in 2014, teams were eligible to buy-back a player using their limited stock of RTM cards for an average of 19.5 players.

During the auctions, teams on average purchased 15 and 11 players, in 2014 and 2011 respectively, comprising approximately of 4-3 batsmen, 6-5 bowlers, 2-1 wicket-keepers and 3-2 all-rounders.²⁶ In 2011, 4 Indian players were bought (on average) in contrast to 10 Indian players purchased in 2014, reflecting the tightening of the overseas quota in 2014. Two additional constraints in 2014, the requirement that teams purchase newcomers and the use of the RTM option meant that teams purchased 6.5 newcomers on average and were able to buy back almost 2 players using the RTM option. Overall these summary statistics suggests that the organizer-imposed team composition constraints are binding and modify how teams value players.

3.2.2 Sequencing of the auction: Order of Sale

At the time of the auction, the auctioneer proceeds with the sale of players according to the predetermined sequence of the sets into which they are categorized. Players are arranged and sequenced into different 'sets' by their cricketing speciality, popularity²⁷ and, to some extent, their reserve price. The composition of the sets and the sequence in which different sets will be placed in the auction are announced ex ante.²⁸ Within each set, the order in which players are presented is determined by random draws in the format of a lottery. The lottery proceeds as follows: the auctioneer picks a chit from an urn that contains the name of the player to be sold first in the set. He then proceeds to auction this player and upon conclusion of the sale, returns to the urn to draw another chit for the subsequent player to be auctioned. The procedure is repeated until the last player has been sold. This randomly selected order of sale acts as an important ingredient for our econometric approach.

²⁶We abstract from the accelerated auctions which are relatively marginal in terms of additional purchases (1.2 and 3.6 players in 2011 and 2014, respectively).

²⁷By popularity, we refer to the ability of a player to attract sponsorship to the team. As a large part of team revenue is sourced from sponsorship, a player's 'marketability' plays an important factor in their selection. See a report by American Appraisal (<http://american-appraisal.co.in/>) for an analysis of teams' 'brand-value' from sponsorship in the IPL.

²⁸Teams are issued a list of the players that are to be auctioned, the composition of the sets and the order in which the sets will be auctioned are determined before the auction.

To illustrate the effect of the order of sale on wages, we present some descriptive evidence from the auctions held in 2014. Figure 1 depicts the reserve price and the final selling price (i.e. the wage when the player is sold) as a function of the overall order of sale in the auction sequence (from 1 to 317). Note that there is a huge heterogeneity in wages. We find that the ratio between the highest and the lowest wage obtained in the auction is as large as 140. We see a strong negative correlation between both, reserve price, as well as final price and order of sale. The correlation is stronger between the final price received and order of sale. Next we provide some descriptive evidence to gauge how the *within-set* order of sale effected the final sale price of a player. Table 4 reports estimates from a reduced-form regression of the within-set order and the square of order (order sq.) on the final sale price of all sold players in the auction. On the whole, the picture is consistent with the empirical literature on sequential auctions: on the one hand, players of higher quality are auctioned first through the set ordering; on the other hand, ceteris paribus there is a declining price trend (more precisely the within-set order had a negative convex effect on the final price of a player).

3.2.3 Auction procedure

The auctioneer opens the auction for a given player at the reserve price and raises the price according to a predetermined increment schedule.²⁹ Starting from a given reservation wage, teams are invited to raise their paddle to indicate their willingness to buy the player at the current price. A team that raises the paddle, when the current price is w , is considered to have the standing bid and can only be displaced by a higher bid ($w + \Delta(w)$, where $\Delta(w)$ is the increment) from a competing team.³⁰ If no competing bidder challenges the standing bid within a given time frame, the team with the standing bid becomes the winner. If the player is not RTM-eligible, then, the player is sold to this team at a price equal to its bid. If the player is RTM-eligible, his team from the previous year, has the option to use one of their RTM cards and match the winning offer to buy-back their player. Our data contain the entire bidding dynamics in each auction (i.e. which team raise the paddle) and the RTM activity if any. In our data, we find that all teams have used their RTM option at least once. Overall, 11% of the players have been purchased through the RTM option.³¹

²⁹The rule for fixing the increments is published ex ante by the auctioneer. The increments are not constant throughout the auction but depend on the current bid. For 2014, the grid (in Millions of rupees) was: $\Delta(w) = 0.5$ if $w < 10$, $\Delta(w) = 1$ if $10 \leq w < 20$, $\Delta(w) = 2$ if $20 \leq w < 25$, $\Delta(w) = 2.5$ if $25 \leq w < 50$, $\Delta(w) = 5$ if $50 \leq w < 125$ and $\Delta(w) = 10$ if $w \geq 125$. For 2011, the grid (in ten thousands of dollars) was: $\Delta(w) = 0.5$ if $w < 10$, $\Delta(w) = 1$ if $10 \leq w < 30$, $\Delta(w) = 2.5$ if $30 \leq w < 50$, $\Delta(w) = 5$ if $50 \leq w < 100$, and $\Delta(w) = 10$ if $w \geq 100$.

³⁰Bidders raise their paddle in a somewhat ad hoc way. In particular, it is not clear what are the rules in the case where two bidders raise their paddle simultaneously, i.e. how ties are resolved. An important aspect consistent with our model is that jump bids are precluded. Bidders are also allowed to reenter the auction after a period of inactivity.

³¹Note a team is free to bid in the auction as the other bidders, even if it holds the RTM option for that player. This is verified in our data where we find some teams that actively participated in bidding even when it held the

We present summary statistics of the auction and performance data in Table 3.³² The average number of bidders participating in an auction for any given player is 2.2 in 2014 and 2.6 in 2011. On average, a bidder’s participation rate is approximately 11% across these auctions and is relatively homogenous amongst bidders.³³ The lower panel of the table provides summary statistics on the performance measure of all sold players, observed from all matches played. We describe the construction of the performance measure in Appendix A.4.

4 The empirical specification

In the following two subsections we present the empirical specification of our performance and pseudo-valuation equations and define the explanatory variables included in these models. How our estimation method can be adapted to take into account the specific rules of the IPL auctions (with the RTM cards) is relegated to the Appendix.

4.1 Specification of the performance equation

For each sold player i performing in team f_i^w , we observe multiple performance measures. Specifically, for each match m where i performs we observe the performance index defined in the Appendix. This match-specific performance measure is denoted $y_{i,f_i^w,m}$. Its specification is similar to the one given in Section 2 except that we now add match-specific dummies to account for the possibility that players perform differently throughout the competition (for instance, at the end of the tournament stakes are higher and players may then play differently). We also add a dummy indicating whether a team bats first or second. For notational simplicity β_m stands simultaneously for both types of dummies. Using obvious notations, the performance equation that we estimate becomes

$$y_{i,f_i^w,m} = \beta_{f_i^w} + \beta_m + h(w_i) + \beta_x \cdot x_{i,f_i^w} + \gamma \cdot CF_{i,f_i^w,m}^{\hat{\alpha}} + \chi \cdot \sum_{f \neq f_i^w} CF_{i,f,m}^{\hat{\alpha}} + \text{error}_{i,f_i^w,m}. \quad (13)$$

The wage appearing in this equation is defined in logarithms. The vector of covariates $x_{i,f}$ includes the following variables: set-fixed effects (that control for cricket-specialities and the fact that especially the early sets contain the more able players); a dummy for whether player i is RTM-eligible (i.e., one of the teams has the possibility to purchase i by using a RTM card); a dummy indicating whether team f is eligible to use a RTM card to purchase i ; a dummy

RTM option for that player.

³²Each team is required to play 14 matches before the play-offs stage. We observe missing performance indicators for some players, if their team fails to reach the play-offs or if they do not get an opportunity to bat or bowl. This typically occurs for a batsman, when all the other batsmen above him remain un-dismissed and the time limit (120 balls) has been reached.

³³The participation rate in 2014 (among the subsample of sold players) of the teams CSK, DC, Dehli, KKR, KXI, Mumbai, RCB, RR are respectively 10%, 13%, 16%, 8%, 8%, 11%, 11% and 11%.

indicating whether the player is of Indian nationality; and finally a dummy called newcomer indicating whether the player has already been called by his national team.³⁴

4.2 Specification of the pseudo-valuation equation

We specify the functions G_f , $f = 1, \dots, F$, as linear functions of the explanatory variables and the parameters:

$$G_f(z_{i,f}, x_{i,f}) = \alpha_f + \alpha_z \cdot z_{i,f} + \alpha_x \cdot x_{i,f}, \quad (14)$$

where α_f is a bidder-specific fixed effect, and α_z and α_x are parameters to be estimated. The errors $\epsilon_{i,f}$ appearing in the pseudo-valuation equation are i.i.d. normally distributed random variables with mean zero and variance α_σ^2 . Recall that the independence assumption on the error terms is required for the non-parametric identification of the auction model. The hypothesis that the variance is invariant across f is not required but is assumed here for simplicity.

The vector of auction variables $z_{i,f}$ includes variables that are assumed to influence the way teams bid in the auction but not the subsequent performance of players. We take: the order of sale of player i within the set and the square of this variable; the budget left to team f just prior to the auction of i ; and finally, for each cricket speciality s , the interaction between a variable counting the number of players of type s already bought by f , and a dummy indicating whether i is of speciality s (since there are four specialities we have four such interaction variables).³⁵

Let us comment on what we should expect regarding the sign of the parameters α_x and α_z . First consider the predicted effects of the variables in x . Given the overseas player quota, there should be a high demand for Indian players, so we expect the dummy for Indian nationality to be positive. The newcomer indicator is a proxy for past cricket experience, and hence we expect this variable to have a positive effect in the pseudo-valuation equation as well. There are no clear predictions on the signs of the two RTM related variables. On the one hand, the presence of a RTM seems a good signal since it reflects experience in the IPL if the player is eligible to RTM and also experience in the specific team considered if this bidder is eligible to RTM. On the other hand, the fact that a given player is not retained could be a signal that the team is no longer (or less) interested in this player or that she prefers to use the auction to fix its wage.

Next consider the effect of the auction variables z . The coefficients on our four backlog variables are expected to take negative signs: when a team already possesses players of a given

³⁴Players who have played in their national teams are actually experienced. We nonetheless use the terminology “Newcomer” as this is the official designation for such players.

³⁵The idea to account for past bidding behavior is reminiscent of [Jofre-Bonet and Pesendorfer \(2003\)](#). In their analysis of repeated highway procurement auctions, the distribution function of bids depends on a variable called backlog, a monetary measure of the amount of work left on previous projects won by the bidder divided by an estimation of the capacity (or size) of the firm. In the dynamic auction literature variables measuring past bidding behavior are nowadays often referred to as backlog variables.

speciality, it should lower the value of buying yet another player of this type. We expect this effect to be strongest for wicket-keepers and batsmen (since teams need relatively few of these types of players to form a well functioning squad). Concerning the effect of the budget, we expect that the more money a team has to spend, the higher it is prone to bid. Predicting the impact of the order of sale is less straightforward. Successive winning prices are, in theory, expected to vary across the auction sequence due to strategic motives, once we depart from the most simple model of sequential auctions with independent private values and symmetric risk-neutral unit-demand bidders.³⁶ The sign of the price trend is either ambiguous or it depends on the specific model assumptions. The empirical literature often reports declining price patterns. While the largest part of this declining price trend comes usually from a strategic ordering of heterogeneous items by auctioneers where the most valuable lots are sold first (see for instance [Beggs and Graddy \(1997\)](#)), such “declining price anomalies” have been also reported for sales of homogeneous goods (see [Ashenfelter \(1989\)](#) and [van den Berg et al. 2001](#)).^{37,38} The implication of all this is that we may anticipate the coefficients on order and order squared to be different from zero, but without knowing their precise signs.

5 Estimation results

We now present results for the effect of wages on performance from our estimation strategy that controls for sample selection and endogeneity. [Table 5](#) reports results from the first stage of our estimation, i.e., the maximum likelihood estimates of the parameters α appearing in the likelihood function [\(11\)](#). Estimations are presented separately for the 317 players in 2014 and the 333 players in 2011. Reported are the ML estimates of α_x and α_z together with the asymptotic standard errors in parentheses.³⁹

Let us first look at the estimated effects of our auction variables. The table shows that the or-

³⁶These assumptions guarantee that prices follow a martingale (see [Milgrom and Weber 2000](#)).

³⁷Some institutional features and/or special ingredients from the auction design may also be at the source of declining prices. [Ginsburgh \(1998\)](#) analyzes the effect of absentee bidders who do not show up physically in the auction house and submit the same bid order throughout the sequence. [Black and de Meza \(1992\)](#) and [Février, Linnemer, and Visser \(2007\)](#) consider auctions where the winner of an early auction wins also the option to buy further units at his winning price. [Ashenfelter and Genesove \(1992\)](#) consider condominium auctions where early winners benefit from a right-to-choose clause. None of these elements are present in the IPL auctions.

³⁸[Deltas and Kosmopoulou \(2004\)](#) show that the influence of the sequence order on the final price distribution can be more subtle by affecting the variance rather than the mean.

³⁹The likelihood [\(11\)](#) corresponds to the English auction of [Section 2](#) where the winner pays the highest bid among his opponents. In the IPL auctions, however, winners pay their own bid. This implies that the English auctions observed in the data have a “first-price auction element” wherein bidders have an incentive to shade their bids. Accounting for this aspect of the IPL pricing rule would require further developments along the lines of the seminal work of [Guerre, Perrigne, and Vuong \(2000\)](#) where the “bid shading” function is estimated first. Another radically different approach to model the English auction with increments (and wherein winners pay their own bids) would be to specify the auction model only partially, i.e., to develop an incomplete model as in [Haile and Tamer \(2003\)](#)). A drawback of this approach is that it prevents identification and provides only bounds on the parameters of interests.

der in which players are auctioned (within set) plays an important role in determining the pseudo-valuations for each player, both in 2011 and 2014. For 2011, we find that order and order squared are strongly significant, and the estimates imply a concave relationship between a player's pseudo-valuation and order of sale. More precisely, up to the seventh player ($0.337/0.048=7.02$) auctioned within a set, the valuations $V_{i,f}$ are increasing in the order of sale.⁴⁰ For example, the pseudo-valuation of the player auctioned second in a set exceeds by 30% the one of the first player in the same set.⁴¹ For 2014, we find that only order squared is significant. The sign is positive so here as well the implication is that the pseudo-valuations are increasing in the order of sale.

We find important effects of the other auction variables as well. For both auction years, the logarithm of the remaining budget is strongly significant: teams bid more aggressively when they have more budget; a 1% increase in a team's remaining budget at any point in the auction, increases a player's pseudo-valuation by 2.3% in 2014 and by slightly more than 1% in 2011. The four backlog variables are significant and the coefficient signs are negative for the 2014 auction: teams are likely to bid less aggressively if they have already bought similar speciality players before. For instance, when a team has already purchased a batsman, it reduces the pseudo-valuation for the additional batsman by 28%. The effect is largest for the player speciality of wicketkeeping, as expected, since teams require only one wicketkeeper in their playing squad (but are allowed to keep some reserves): a team which has already acquired a wicketkeeper reduces its pseudo-valuation for an additional such type of player by 67%. The backlog variables do not appear to be significant in 2011. More generally the differences in first-stage estimates between 2011 and 2014 seem puzzling at first glance. There is however an important difference between the 2011 and 2014 auctions: in 2014, the auctions were the only way to purchase players while in 2011 half of the purchases were done outside the auctions. The continuation values are thus completely different between 2011 and 2014, so that there is no reason that the auction variables should impact pseudo-valuations in the same way.

Regarding the player-team characteristics (the variables in x), we find large significant effects of both our RTM related variables. A player that can be potentially bought with an RTM card sees an increase in his pseudo-valuation by 63%. A team which is allowed to purchase a player using the RTM option augments its valuation for this player valuation by 540%. The variable newcomer is also significant, and its effect is positive, which is as expected as this variable captures past cricket experience. We also find the expected effect of the overseas quota: although not significant, teams increase their pseudo-valuation for an Indian player by 47% in 2011 and by 72% in 2014.

Next we turn to the results from the second stage of our estimation. Table 6 reports the second

⁴⁰On average the number of players auctioned per set is 10.09 in 2011 and 9.32 in 2014. In the representative set the pseudo-valuations are thus increasing for the first seven players, and declining for the last two or three ones.

⁴¹In calculating this change we use that the left-hand side of the pseudo-valuation equation is defined in logarithms (see equation (2)).

stage estimates for 2014. The dependent variable for all specifications in this table is the match specific batting quartile, i.e., the performance quartile assigned to each player based on his batting performance in the match (4 being the highest and 1 being the lowest). Recall that the player's batting measure is defined as an index based on his score and the speed with which he obtains that score (see appendix A.4 for details). The function h is specified as a parameter multiplied by the wage (columns 1-4); in column 5 we include in addition a second parameter times the interaction between the wage and the week of the tournament, the latter variable running from 1 to 7. First, in columns 1 and 2, we show results from a simple OLS estimation of wages on performance, with and without set fixed effects respectively. The results indicate that player wage has a positive and significant effect on his performance in a match. A 10% increase in wages is associated with a performance increase of 0.025 quartiles (column 2).

Columns 3-5 reports selection corrected results. The standard errors for all selection corrected results are obtained by a non-parametric block-bootstrap method which is clustered at the player level. In column 3, we include the control function terms for both the winning (γ) and losing (χ) bidders. Using our control function approach, we find that the effect of wages is no longer significant. The effect of the correction term for the winner is strongly significant, but the effect of the term for the losers is not. Since in addition $\gamma > 0$ there is a positive correlation between the error terms u and ϵ for the auction winner. The comparison between the coefficients of the winner and loser correction terms, specifically that $\gamma > \chi = 0$, indicate that the auction reflects a private value rather than a common value paradigm.

Since we cannot reject the null that $\chi = 0$, we drop the control function term for the losing bidders and retain only that for the winner, to improve precision. This leads to a tighter confidence interval for γ , as seen in column 4. Finally, column 5 adds the interaction of the wage and the week of the tournament. In this specification, the coefficient on this interaction term is significant (while the coefficient on wage remains insignificant), implying that the wage effect increases over time. A 10% increase in wages increases performance by 0.048 quartiles in the last (seventh) week of the tournament.

Table 7 present similar results for two alternative performance measures, a raw batting performance measure (batsman strike rate) (columns 1-2) and an overall (batting and bowling) performance quartile (columns 3-4). Both measures are defined in detailed in Appendix A.4. The latter measure, reflects the overall performance of a player on a variety of dimensions, in contrast to the previous measure, which is entirely batsman specific. Despite this advantage, it is subject to more noise given the difficulty in comparing speciality based performances, affecting the precision of our estimates. The pattern of the results are largely unchanged. The coefficient γ remains positive and is statistically significant in columns 1-2 but loses significance in columns 3-4, possibly due to the noisiness of the dependent variable in these specifications. As before we find that the overall wage effect is insignificant after controlling for selection but find a positive and significant impact of wage interacted with week.

Next, we analyze wage effects for 2011. The important difference between player auctions in 2014 and 2011 was with respect to the nature of player contracts. As explained earlier, in 2014 players were offered three year renewable wage contracts, that could be terminated at any time at the discretion of the hiring team. In contrast, in 2011, players were offered a fixed three year contract, which implicitly prevented teams from firing any hired player through the three year term. A priori, the fixed-term nature of the wage contract offered no variable incentive for the player to perform well, especially toward the end of the tournament, as there was no pre-announced secondary auction. We therefore exploit the difference in wage contracts between the two auction years to test whether the performance based incentive explanation holds true.

Table 8 presents the second stage results for 2011. The dependent variable for all specifications here is, like in Table 6, the match specific batting quartile. As before, we find a positive and significant effect of wages on player performance from a simple OLS (column 1) which does not account for selection (although the impact loses significance when set fixed effects are added to the specification, as shown in column 2). A 10% increase in wages increases performance by 0.013 quartiles. However, this positive wage effect disappears, once we account for selection (columns 3-4). The selection correction term for the winner is significant, indicating that the positive wage effect previously found is a result of omitted player characteristics which are related to wages and sample selection. More importantly, after accounting for selection, we find no heterogeneous effects of wages on performance by the week of the tournament (column 4). This result sharply contrasts with the results obtained for 2014 where we found strong positive and significant wage effects towards the end of the tournament.

The difference in the wage by week effects (i.e., the effect for the interaction term between the week of the tournament and wage) between 2011 and 2014, can potentially be explained by the fact that players were offered renewable contracts in 2014. With renewable contracts teams were allowed to terminate the employment of under-performing players, who would then be re-allocated (or be unemployed) to other teams through a secondary auction at the end of the tournament. The possibility of being fired presented an important incentive for players to perform well in the tournament. In addition, it could be the case that these unemployment or wage loss concerns could be more binding for players with a high wage, since they stand to lose more (compared to low wage players), if their teams decide to re-auction them (see [Krautmann and Solow \(2009\)](#) for a related explanation). This can explain why the interaction term between the week of the tournament and wage has a positive and significant effect on player performance /effort in 2014 compared to 2011 where we find no such effect.

Finally in table 9 we test whether wage effects are driven by fairness considerations. We focus on the results for 2014 (those of 2011 are similar). The dependent variable for all specifications in this table is the match specific batting quartile again. We analyze whether players differentially respond to an increase in wages with respect to their reference wage. The reference wage, w_i^{ref} , is calculated as the mean wage among players in a reference group. To examine possible

fairness effects the function h is specified as: $h(w_i) = \beta_w \cdot w_i + \beta_w^- \cdot w_i \cdot 1\{w_i < w_i^{ref}\} + \beta_r \cdot 1\{w_i < w_i^{ref}\}$, where β_w^- is the coefficient on the wage of a below-mean (wage) player and β_r is the coefficient for being a below-mean (wage) player. Various definitions of the reference group are considered. In column 1 the player’s reference group consists of all players with a similar reserve price, column 2 all players of the same team, column 3 all players sold in the same set, column 4 the full sample of sold players, column 5 all players of the same speciality, and column 6 all players of the same nationality.⁴² The OLS results (presented in the upper panel) indicate that wage generally has a positive and significant effect, regardless of the type of reference group. The coefficients β_w^- and β_r are, however, not significant. The corrected estimates (lower panel) indicate that the wage is no longer significant, and for all types of reference groups the control function term (for the winner) is significant. Even after correcting for sample selection and endogeneity, there is no evidence of any fairness effects.

6 Conclusion

In this paper we propose a method to estimate the relationship between the price of a good sold at auction, and a post-auction outcome which is observed among auction winners. We develop estimation methods to analyze such relationships, because typically, both, the allocation of items to bidders and the price of the item, is endogenously determined. This induces a bias in causally identifying the impact of the auction determined price on some post-auction outcome. Our proposed method corrects for these multiple sources of bias via a control function approach based on the non-parametric identification of an auction model.

Our empirical procedure, applied to a setting examining the wage-performance relationship, has two steps: First, we model the auction behavior of firms (who buy workers) under the assumption that the workers are sold one after the other, through sequential English auctions with bid increments. The control function is the expectation of the error term appearing in the performance equation conditional on the full information set observed by the econometrician. We show that the auction model and hence the control function are non-parametrically identified. Second, the control function is added to the performance equation, and the augmented equation is estimated by standard regression techniques.

We then apply our methodology to unique field data from auctions of cricket players and their game-specific performances during an important tournament held in India. The winning bids in these auctions represent the salaries the auction winners (the teams) have to pay to the players for their participation in the tournament. Our empirical strategy exploits the fact that several features of the auctions act as exogenous shifters of wages. Our results indicate that the naive OLS estimate of the wage effect is statistically significant and positive. This effect disappears

⁴²In all cases, we include i in the reference group, so that w_i^{ref} is always defined.

once we properly account for sample selection and endogeneity using our control function terms derived from our first stage auction model. These terms are throughout significant and have positive effects, indicating a positive correlation between the error terms appearing in the performance and pseudo-valuation equations. For the auctions of 2014 we find a positive effect of the wage interacted with a week indicator, suggesting that the wage impact increases over time when players face the threat of being reaucted after the first tournament season. Finally, using different definitions of reference groups, we find no evidence of fairness in the data.

Our methodological approach consists broadly in using the econometrics of auctions to build control functions in environments where selection comes from competitive bidding. It can potentially be applied to other settings where one is interested in the relationship between a post-auction outcome and the auction price of a good or service. Other auction environments are allowed as well, provided that the auction model is identified so that we can compute the control function $E(u_{i,f,w}|\mathcal{I})$ and thus control for the endogeneity of the auction price and the sample selection. We provide a brief discussion of some possible alternative application areas below.

Banking and Macro-finance: There are many studies that seek to understand how treasury auctions impact yields and spread as well as secondary market price movements. For example, [Joyce and Tong \(2012\)](#) study the impact of an increase in bond supply, through a quantitative easing program, on post-auction bond yields. They regress percentage point change in each bond's yield over the day of each auction on the amount purchased of that bond. Treasury bonds are sold/repurchased by the central bank through multi-unit auctions. Bidders in these auctions, usually private banks, are required to submit multiple bids, consisting of both price and quantity. Since both quantities and prices of bonds are determined through the auction determined competitive bidding process, analyzing post-auction outcomes will be typically subject to both endogeneity and selection concerns. In a pure private value environment as e.g. in [Hortaçsu and Puller \(2008\)](#) or [Kastl \(2011\)](#) for the uniform price auction,⁴³ a structural bidding model allows to estimate the distribution of bidders' valuations given any auction outcome. In a related vein, [Cassola, Hortaçsu, and Kastl \(2013\)](#) shows that the regression of a bank's performance/profitability on auction-based measures supposed to be proxies of a bank's short-term funding costs depends crucially whether the auction-based measure is the final bid (a reduced form approach) or an estimation of the bank's willingness to pay (estimated by a structural approach).

Corporate Finance: A large literature in empirical corporate finance has examined the post-acquisition outcomes of companies when they are acquired through an auction procedure (see

⁴³[Kastl \(2011\)](#) considers that each bidder submits a step function instead of a continuous function as in [Hortaçsu and Puller \(2008\)](#). Such a structural analysis can be easily adapted to the discriminatory auction which is often used in treasury auctions.

Eckbo (2009) for a review).⁴⁴ An issue for bankruptcy auction is whether recovery rates are higher when the previous owner of the bankrupt firm ‘wins’ the auction (Thorburn 2000). The endogeneity and sample selection problems arise because previous owners tend to repurchase the firm when they have private information on the quality of the firm and also because recovery rates are only observed for firms who have been successful in selling their assets.

⁴⁴Malmendier, Moretti, and Peters (2012) consider competition between firms to merger with another firm. In general the sale procedure may not be a formal auction but rather a bargaining procedure that is modeled as an English auction.

References

- AKERLOF, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97, 543–569.
- AKERLOF, G. A. AND J. L. YELLEN (1990): "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105, 255–283.
- ASHENFELTER, O. (1989): "How Auctions work for Wine and Art," *Journal of Economic Perspectives*, 3, 23–36.
- ASHENFELTER, O. AND D. GENESOVE (1992): "Testing for Price Anomalies in Real-Estate Auctions," *American Economic Review*, 82, 501–05.
- ATHEY, S. AND P. HAILE (2002): "Identification of Standard Auction Models," *Econometrica*, 70, 2107–2140.
- (2008): *Nonparametric Approaches to Auctions*, Amsterdam: North-Holland, Handbook of Econometrics, Vol. 6.
- BALDWIN, L. H., R. C. MARSHALL, AND J.-F. RICHARD (1997): "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy*, 105, 657–99.
- BEGGS, A. AND K. GRADDY (1997): "Declining Values and the Afternoon Effect: Evidence from Art Auctions," *RAND Journal of Economics*, 28, 544–565.
- BIKHCHANDANI, S., S. LIPPMAN, AND R. RYAN (2005): "On the Right of First Refusal," *The B.E. Journal of Theoretical Economics*, 5, 1–44.
- BLACK, J. AND D. DE MEZA (1992): "Systematic Price Differences Between Successive Auctions are no Anomaly," *Journal of Economics and Management Strategy*, 1, 607–628.
- BLUNDELL, R., A. DUNCAN, AND C. MEGHIR (1998): "Estimating labor supply responses using tax reforms," *Econometrica*, 827–861.
- CASSOLA, N., A. HORTAÇSU, AND J. KASTL (2013): "The 2007 Subprime Market Crisis Through the Lens of European Central Bank Auctions for Short-Term Funds," *Econometrica*, 81, 1309–1345.
- COHN, A., E. FEHR, AND L. GOETTE (2015): "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment," *Management Science*, 61, 1777–94.
- DAS, M., W. NEWEY, AND F. VELLA (2003): "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58.
- DELTAS, G. AND G. KOSMOPOULOU (2004): "'Catalogue' vs 'Order-of-sale' effects in sequential auctions: theory and evidence from a rare book sale," *Economic Journal*, 114, 28–54.

- DUBIN, J. A. AND D. L. MCFADDEN (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–62.
- ECKBO, B. E. (2009): "Bidding strategies and takeover premiums: A review," *Journal of Corporate Finance*, 15, 149–178.
- FEHR, E., G. KIRCHSTEIGER, AND A. RIEDL (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108, 437–59.
- FÉVRIER, P., L. LINNEMER, AND M. VISSER (2007): "Buy or wait, that is the option: the buyer's option in sequential laboratory auctions," *RAND Journal of Economics*, 38, 98–118.
- GINSBURGH, V. (1998): "Absentee Bidders and the Declining Price Anomaly in Wine Auctions," *Journal of Political Economy*, 106, 1302–1319.
- GNEEZY, U. AND J. A. LIST (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74, 1365–1384.
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.
- HAILE, P. AND E. TAMER (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.
- HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–61.
- HENDRICKS, K., J. PINKSE, AND R. PORTER (2003): "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions," *Review of Economic Studies*, 70, 115–145.
- HICKMAN, B. (2010): "On the pricing rule in electronic auctions." *International Journal of Industrial Organization*, 28, 423–433.
- HORTAÇSU, A. AND S. L. PULLER (2008): "Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market," *The RAND Journal of Economics*, 39, 86–114.
- IMBENS, G. AND J. WOOLDRIDGE (2009): "Control function and related methods," CEMMAP Lecture Notes 14.
- JOFRE-BONET, M. AND M. PESENDORFER (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489.
- JOYCE, M. A. AND M. TONG (2012): "QE and the Gilt Market: a Disaggregated Analysis*," *The Economic Journal*, 122, 348–384.

- KASTL, J. (2011): "Discrete Bids and Empirical Inference in Divisible Good Auctions," *Review of Economic Studies*, 78, 974–1014.
- KRAUTMANN, A. C. AND J. L. SOLOW (2009): "The dynamics of performance over the duration of Major League Baseball long-term contracts," *Journal of Sports Economics*, 10, 6–22.
- LAFFONT, J.-J. AND Q. VUONG (1996): "Structural Analysis of Auction Data," *American Economic Review (Papers and Proceedings)*, 86, 414–20.
- LEE, D. AND N. G. RUPP (2007): "Retracting a Gift: How Does Employee Effort Respond to Wage Reductions?" *Journal of Labor Economics*, 25, 725–761.
- MALMENDIER, U., E. MORETTI, AND F. PETERS (2012): "Winning by Losing: Evidence on Overbidding in Mergers," *Unpublished Manuscript*.
- MAS, A. (2006): "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121, 783–821.
- MEILIJSON, I. (1981): "Estimation of the Lifetime Distribution of the Parts from the Autopsy Statistics of the Machine," *Journal of Applied Probability*, 18, 829–838.
- MILGROM, P. AND R. J. WEBER (2000): *A Theory of Auctions and Competitive Bidding, II*, PKlempere. Cheltenham: Edward Elgar Publishing, Ltd, vol. 2, 179–194.
- PAARSCH, H. J. (1997): "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbian Timber Sales," *Journal of Econometrics*, 78, 333–357.
- PAARSCH, H. J. AND H. HONG (2006): *An Introduction to the Structural Econometrics of Auction Data*, MIT Press.
- THORBURN, K. S. (2000): "Bankruptcy auctions: costs, debt recovery, and firm survival," *Journal of financial economics*, 58, 337–368.
- VAN DEN BERG, G. J., J. C. VAN OURS, AND M. P. PRADHAN (2001): "The Declining Price Anomaly in Dutch Dutch Rose Auctions," *Amer. Econ. Rev.*, 91, 1055–1062.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Table 1: Auction Rules

	Auction 2011	Auction 2014
# of Teams	10	8
# of Players Auctioned	350	514
Player Contract	Three year fixed-term contract	One year contract renewable for two more years.
Retain Players	Maximum of 4	Maximum of 5
Spending caps	4 Retained Players = \$4.5 million 3 Retained Players = \$5 million 2 Retained Players = \$5.9 million 1 Retained Players = \$7.2 million 0 Retained Players = \$9 million	5 Retained Players = \$3.5 million 4 Retained Players = \$4.1 million 3 Retained Players = \$5 million 2 Retained Players = \$6.3 million 0 Retained Players = \$10 million
Overseas	Not more than 10 overseas players	Not more than 9 overseas players
Right to Match (RTM)	None	5-3 Retained Players = 1 RTM 1-2 Retained Players = 2 RTM 0 Retained Players = 3 RTM

Table 2: Bidder summary statistics

Variable	Description	2014 (8 bidders)		2011 (10 bidders)	
		Mean	Std. Dev.	Mean	Std. Dev.
<i>Bidder purchases in the auction:</i>					
# Players	Number of players bought in the auction	15.25	2.49	11.1	3.51
speciality: # Batsman	Number of batsmen bought in the auction	3.75	1.98	3.4	1.89
speciality: # Bowler	Number of bowlers bought in the auction	6.5	2.07	5	1.76
speciality: # Wicket-Keeper	Number of wicket-keepers bought in the auction	1.62	0.74	0.9	0.73
speciality: # All-Rounder	Number of all-rounders bought in the auction	3.37	1.59	1.8	1.39
Nationality: # Indian	Number of Indian players bought in the auction	10.12	1.72	4.3	1.63
# Newcomers	Number of players bought in the auction who are newcomers	6.5	1.60	-	-
# RTM Bought Players	Number of players bought in the auction through RTM	1.62	0.744	-	-
<i>Bidder constraints in the auction</i>					
# Retained Players	Number of players retained by teams before the auction	3	1.85	1.2	1.62
# RTM eligible Players	Number of players eligible to be bought back using RTM (per team)	19.5	4.14	-	-
Spending cap	Amount of money allocated to a team net of retained players	5.01	1.90	7.23	2.15
Remaining budget	Unused budget of a team at the end of the auction	0.53	9.38	1.28	1.10

Note: All budget figures 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 70 (INR).

Table 3: Summary statistics on auction and performance data

Auction data

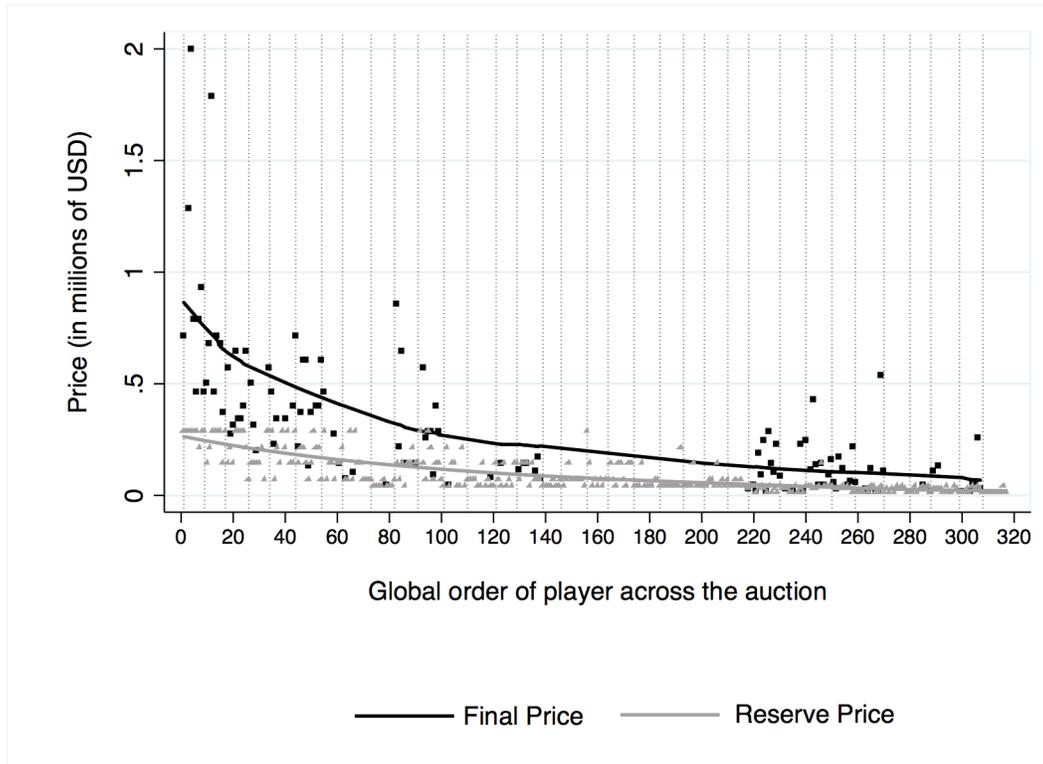
Variable	Description	2014		2011	
		Mean	Std. Dev.	Mean	Std. Dev.
<u>Full sample:</u> N=317					
Reserve price Order	Reservation wage set by auctioneer	0.09	0.08	0.11	0.10
Indian	Within-set order of player appearance in auction	5.23	2.80	5.65	3.11
Newcomer	Dummy indicating whether player is Indian	0.44	0.50	0.14	0.35
RTM	Dummy indicating whether player is a newcomer	0.32	0.47	-	-
	Dummy indicating whether player is eligible for RTM	0.50	0.50	-	-
<u>Sold players sample:</u> N=122					
# active bidders	Participating bidders for each player auction	2.28	1.08	2.62	1.29
Winning price	Equal to the final wage of the player	0.29	0.32	0.54	0.50
Reserve price	Reservation wage set by auctioneer	0.11	0.10	6.03	2.81
Order	Within-set order of player appearance in auction	5.29	2.90	1	12
Indian	Dummy indicating whether player is Indian	0.66	0.47	0.39	0.49
Newcomer	Dummy indicating whether player is a newcomer	0.43	0.50	-	-
RTM	Dummy indicating whether player is eligible for RTM	0.75	0.44	-	-

Performance data

Variable	Description	2014		2011	
		Mean	Std. Dev.	Mean	Std. Dev.
<u>Batsmen sample:</u> N=550					
Batting strike rate	Average score of a batsman per 100 balls faced	114.12	70.88	103.54	66.32
Batting quartile	Quartile of player based on his batting performance index	2.38	1.10	2.41	1.14
<u>All players sample:</u> N=859					
Performance quartile	Quartile of player based on his batting and bowling performance index	2.57	1.07	2.690	1.09

Note: 1. Performance data are summarized across matches for players that are sold in the auction and perform in a given match; 3. † Construction of batting and bowling index is described in Appendix;

Figure 1: Order of Sale and Final/Reserve Price of All Players, 2014



This figure plots the final price (for a sample of sold player) and the reserve price (for all players) against the order of sale in the auction. The dotted vertical lines in gray indicate each different 'set' in the auction.

Table 4: Within-Set Effect of Order on Final Price of Sold Players, 2014

	(1)	(2)	(3)
Order	-0.177 (0.112)	-0.198* (0.100)	-0.206** (0.104)
Order Sq.	0.017* (0.010)	0.019** (0.009)	0.020** (0.009)
Control for Reserve Price	No	Yes	Yes
Control for Player Attributes	No	No	Yes
Set Fixed Effects	Yes	Yes	Yes
<i>N</i>	122	122	122

Note: This table reports coefficient estimates and standard errors (in parentheses) of the order and order-square when regressed upon the log of the final price. The regression is estimated on a sample of 122 sold players. * indicates significance at 10%; ** at 5%; *** at 1%.

Table 5: First Stage: Pseudo-Valuation Determinants

	2011	2014
<i>Player-team (ex ante) Characteristics:</i>		
Player eligible to RTM		0.486** (0.195)
Bidder eligible to use RTM		1.867*** (0.342)
Newcomer		1.699*** (0.332)
Indian	0.382 (0.398)	0.541 (0.616)
<i>Auction Variables:</i>		
Order	0.337*** (0.074)	-0.184 (0.121)
Order Sq.	-0.024*** (0.006)	0.020* (0.011)
Remaining Budget (in logs)	1.057*** (0.259)	2.342*** (0.312)
# Batsmen bought	0.022 (0.054)	-0.329*** (0.083)
# Bowlers bought	0.057 (0.048)	-0.098* (0.059)
# Wicket-keepers bought	-0.150 (0.102)	-1.103*** (0.260)
# All-Rounders bought	-0.627 (0.508)	-0.198** (0.090)
Constant	-7.644** (3.649)	-29.468*** (5.642)
# Auctioned Players	333	317
# Bidders	10	8

Note: This table reports ML estimates of α_x and α_z and asymptotic standard errors (in parentheses). The auction in 2011 and 2014 consisted of 10 and 8 bidding teams respectively. Maximum likelihood estimation is based on the full sample of auctioned players (333 in 2011 and 317 in 2014). Variable definition for the covariates are provided in Tables (2) and (3). * indicates significance at 10%; ** at 5%; *** at 1%.

Table 6: Second Stage: Effect of Wage on Performance, 2014

	Dep. Variable: Batting Quartile				
	OLS		Selection Corrected		
Wage	0.110*	0.225***	-0.157	0.108	-0.148
	(0.064)	(0.083)	(0.274)	(0.132)	(0.201)
Wage × Week					0.069**
					(0.033)
CF Winner (γ)			0.423***	0.316**	0.325**
			(0.156)	(0.128)	(0.128)
CF Losers (χ)			0.033		
			(0.029)		
Week	0.076	0.069	0.165	0.160	-0.999*
	(0.114)	(0.115)	(0.109)	(0.110)	(0.564)
Set FE	No	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes
# Obs.	550	550	550	550	550

Note: This table reports coefficient estimates and standard errors (in parentheses) from the second stage, i.e., from a regression of player performance on the wage for the year 2014. The regression is estimated on the sample of all matches played by the 122 sold players in 2014. Performance is measured as the match-specific batting quartile of a player, as defined in Appendix A.4. *Wage* refers to the salary (log) of the player, and is equivalent to his final bid price in the auction; *Week* refers to the week of the tournament (ranging from 1-7); *CF Winner* (resp. *CF Losers*) refers to the control function term of the winner of the auction (resp. losers). All specifications account for a full set of fixed effects with respect to the set in which the player was auctioned (*Set FE*), the player's team (*Team FE*) as well as match fixed effects. In addition all specifications control for player attributes (dummy for overseas player, specialty, dummy for newcomer) and match-specific variables (dummy for qualifier, dummy for whether the team batted first or second). Standard errors reported in parentheses are non-parametrically block-bootstrapped and clustered at the player-level. * indicates significance at 10%; ** at 5%; *** at 1%.

Table 7: Second Stage: Effect of Wage on Alternate Performance Measures, 2014

	<u>Batting Strike Rate</u>		<u>Batting+Bowling Quartile</u>	
Wage	0.491 (10.776)	-15.192 (15.368)	0.119 (0.080)	-0.137 (0.115)
Wage × Week		4.223* (2.467)		0.066*** (0.024)
CF Winner (γ)	22.822** (10.353)	23.357** (9.998)	0.106 (0.078)	0.111 (0.079)
Week	36.108** (16.330)	-35.062 (44.618)	0.046 (0.087)	-1.050** (0.410)
Set FE	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes
# Obs.	550	550	917	917

Note: This table reports coefficient estimates and standard errors in parentheses) from the second stage, i.e., from a regression of player performance on the wage for the year 2014. The regression is estimated on the sample of all matches played by the 122 sold players in 2014. Performance is measured as the match-specific batting strike rate of a player in columns (2)-(3) and combined batting and bowling quartile of a player, as defined in Appendix A.4. *Wage* refers to the salary (log) of the player, and is equivalent to his final bid price in the auction; *Week* refers to the week of the tournament (ranging from 1-7); *CF Winner* refers to the control function term of the winner of the auction. All specifications account for a full set of fixed effects with respect to the set in which the player was auctioned (*Set FE*), the player's team (*Team FE*) as well as match fixed effects. In addition all specifications control for player attributes (dummy for overseas player, specialty, dummy for newcomer) and match-specific variables (dummy for qualifier, dummy for whether the team batted first or second). Standard errors reported in parentheses are non-parametrically block-bootstrapped and clustered at the player-level. * indicates significance at 10%; ** at 5%; *** at 1%.

Table 8: Second Stage: Effect of Wage on Performance, 2011

	Dep. Variable: Batting Quartile			
	OLS		Selection Corrected	
Wage	0.133** (0.059)	0.091 (0.114)	-0.199 (0.418)	-0.123 (0.363)
Wage × Week				-0.014 (0.023)
CF Winner (γ)			0.675* (0.415)	0.676* (0.407)
CF Losers (χ)			-0.060 (0.083)	-0.061 (0.076)
Week	0.011 (0.025)	0.002 (0.233)	0.002 (0.017)	0.192 (0.311)
Set FE	No	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes
# Obs.	641	641	641	641

Note: This table reports coefficient estimates and standard errors (in parentheses) from the second stage, i.e., from a regression of player performance on the wage for the year 2011. The regression is estimated on the sample of all matches played by the 111 sold players in 2011. Performance is measured as the match-specific batting quartile of a player, as defined in Appendix A.4. *Wage* refers to the salary (log) of the player, and is equivalent to his final bid price in the auction; *Week* refers to the week of the tournament (ranging from 1-7); *CF Winner* (resp. *CF Losers*) refers to the control function term of the winner of the auction (resp. losers). All specifications account for a full set of fixed effects with respect to the set in which the player was auctioned (*Set FE*), the player's team (*Team FE*) as well as match fixed effects. In addition all specifications control for player attributes (dummy for overseas player, specialty, dummy for newcomer) and match-specific variables (dummy for qualifier, dummy for whether the team batted first or second). Standard errors reported in parentheses are non-parametrically block-bootstrapped and clustered at the player-level. * indicates significance at 10%; ** at 5%; *** at 1%.

Table 9: Fairness Effects, 2014 (Batsman Quartile)

OLS						
	Reserve Price	Team	Set	All	Specialty	Nationality
Wage	0.220 (0.135)	0.204 (0.132)	0.181 (0.134)	0.306** (0.136)	0.234* (0.136)	0.299** (0.131)
(Wage < Ref. Wage) × Wage	0.067 (0.136)	-0.222 (0.195)	0.064 (0.124)	-0.108 (0.230)	-0.165 (0.226)	-0.076 (0.180)
Wage < Ref. Wage	-0.991 (2.311)	3.282 (3.147)	-1.038 (2.126)	1.948 (3.687)	2.619 (3.612)	1.435 (2.936)
Selection Corrected						
	Reserve Price	Team	Set	All	Specialty	Nationality
Wage	0.147 (0.168)	0.122 (0.193)	0.159 (0.179)	0.209 (0.218)	0.137 (0.217)	0.200 (0.189)
(Wage < Ref. Wage) × Wage	0.016 (0.175)	-0.266 (0.244)	-0.048 (0.142)	-0.159 (0.287)	-0.207 (0.276)	-0.110 (0.252)
Wage < Ref. Wage	-0.122 (2.913)	4.026 (3.947)	0.938 (2.415)	2.750 (4.649)	3.286 (4.486)	1.986 (4.118)
CF Winner (γ)	0.265** (0.111)	0.270** (0.108)	0.286** (0.117)	0.268** (0.106)	0.270** (0.108)	0.267** (0.106)
Set FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	550	550	550	550	550	550

Note: This table reports coefficient estimates and standard errors (in parentheses) from a regression testing fairness effects. The regression is estimated on the sample of all matches played by the 122 sold players in 2014. The first panel of the table reports OLS estimates while the second panel reports selection corrected estimates. Each column of the table reports estimates from using a specific reference group (Reserve Price takes as reference group all players with the same (auction) reserve price as the target player; similarly Team, Set, Specialty, Nationality uses the players in the same team, (auction) set, specialty and nationality as reference group respectively. All takes as a (global) reference group all players.). Performance is measured as the match-specific batting quartile of a player, as defined in Appendix A.4. Wage refers to the salary (log) of the player, and is equivalent to his final bid price in the auction; (Wage < Ref. Wage) is a dummy variable indicating whether a player's wage is below the average wage in his reference group; CF Winner refers to the control function term of the winner of the auction). All specifications account for a full set of fixed effects with respect to the set in which the player was auctioned (Set FE), the player's team (Team FE) as well as match fixed effects. In addition all specifications control for player attributes (dummy for overseas player, specialty, dummy for newcomer) and match-specific variables (dummy for qualifier, dummy for whether the team batted first or second). Standard errors reported in parentheses are non-parametrically block-bootstrapped and clustered at the player-level. * indicates significance at 10%; ** at 5%; *** at 1%.

A Appendix

A.1 Proof of Proposition 1

For any given $x \geq 0$, we define by induction the list $\Delta_0^x, \Delta_1^x, \dots, \Delta_n^x, \dots$ by the initialization $\Delta_0^x = x$ and the recursive relation $\Delta_{n+1}^x = \Delta_n^x + \Delta(\Delta_n^x)$. In words, this corresponds to the list of the possible final prices if the reservation value is equal to x . Since we have assumed that $a := \inf_{x \in \mathbb{R}_+} \Delta(x) > 0$, we have $\Delta_n^x \geq x + a \cdot n$ and then Δ_n^x goes to infinity when n goes to infinity.

The argument of the proof is composed of three steps: (1) Fix the vector of covariates (z, x) and take a given reserve price $W^r \geq 0$, then we show that we can identify from the final wage and the identity of the winner the vector of probabilities $(H_f(\log[\Delta_n^{W^r}] - G_f(z_f, x_f)))_{f=1, \dots, F}$ for each $n \in \mathbb{N}$. (2) Once we have fixed the covariates at (z^*, x^*) , thanks to the variation of the reserve price W^r in the interval $[0, \Delta(0)]$, we obtain the non-parametric identification of the distributions $H_f, f = 1, \dots, F$, on $(-\infty, +\infty)$. (3) We show that the functions $G_f(\cdot, \cdot)$ are identified.

(1) Proof of the first part. Let us consider auctions where only two bidders are eligible (we use here the exogenous variations in the set of eligible bidders),⁴⁵ say f and f' , and fix the covariates and the reserve price W^r . What we observe from this subsample of the data are the (conditional) probabilities that firm f and f' wins the auction at price $\Delta_n^{W^r}$ for each $n \in \mathbb{N}$. Let $\pi_{n,j}^{W^r}$ denote the corresponding probability that bidder $j = f, f'$ wins the auction and has to pay price $\Delta_n^{W^r}$. Let $u_{n,j}^{W^r} = H_j(\log[\Delta_n^{W^r}] - G_f(z_{i,j}, x_{i,j}))$ for $j = f, f'$. Our aim below is to show that the lists $(u_{n,j}^{W^r})_{n \in \mathbb{N}}$ are identified for $j = f, f'$.

The proof is by induction.

Initialization step From the observation of the outcome of the auctions where only bidder f (resp. f') is “eligible”, the probability that the good remains unsold corresponds to $u_{0,f}^{W^r}$ (resp. $u_{0,f'}^{W^r}$) for any possible point in the support of the vector of covariates (z, x) . $u_{0,f}^{W^r}$ and $u_{0,f'}^{W^r}$ are thus identified.

Induction step We show next that if we have already identified $u_{n,j}^{W^r}$ for $j = f, f'$ and any $n \leq n^*$, then we can identify $u_{n^*+1,f}^{W^r}$ and $u_{n^*+1,f'}^{W^r}$ from $\pi_{n^*,f}^{W^r}$ and $\pi_{n^*,f'}^{W^r}$.

Case where $n^* > 0$. If $n^* > 0$, f wins the auction at price $\Delta_{n^*}^{W^r}$ if and only if bidder f' has a valuation in the interval $[\Delta_{n^*}^{W^r}, \Delta_{n^*+1}^{W^r})$ and bidder f has either a valuation in the interval $[\Delta_{n^*}^{W^r}, \Delta_{n^*+1}^{W^r})$ and has won the tie or has a valuation above $\Delta_{n^*+1}^{W^r}$. Formally, using the inde-

⁴⁵Instead of considering exogenous variations in eligibility/participation, [Athey and Haile \(2002\)](#) (or equivalently [Athey and Haile \(2008\)](#)) consider exogenous variations in the vector of covariates and such that any subset of bidders can have arbitrary “bad covariates” (formally $G_f(z_f, x_f)$ can take values that are arbitrary small): by exploiting the limiting cases where all bidders except a subset S have “bad” covariates, then it is as if only the bidders in S are eligible.

pendence between the signals ϵ_f and $\epsilon_{f'}$, and since the probability to win a tie with only one competitor is $\frac{1}{2}$, we have:

$$\pi_{n^*,f}^{W^r} = \left(u_{n^*+1,f'}^{W^r} - u_{n^*,f'}^{W^r} \right) \cdot \left(1 - u_{n^*+1,f}^{W^r} + \frac{1}{2}(u_{n^*+1,f}^{W^r} - u_{n^*,f}^{W^r}) \right)$$

and an analog (symmetric) expression for $\pi_{n^*,f'}^{W^r}$.

Let $Z_f := (u_{n^*+1,f'}^{W^r} - u_{n^*,f'}^{W^r}) \cdot (1 - u_{n^*,f}^{W^r})$ and $Z_{f'} := (u_{n^*+1,f}^{W^r} - u_{n^*,f}^{W^r}) \cdot (1 - u_{n^*,f'}^{W^r})$. In order to identify $u_{n^*+1,f}^{W^r}$ and $u_{n^*+1,f'}^{W^r}$, we show equivalently that we identify Z_f and $Z_{f'}$ under the constraint that $Z_f, Z_{f'} \in [0, \eta]$ where $\eta = (1 - u_{n^*,f}^{W^r}) \cdot (1 - u_{n^*,f'}^{W^r})$. We have

$$Z_f - \frac{1}{2} \frac{Z_f Z_{f'}}{\eta} = \pi_{n^*,f}^{W^r} \quad \text{and} \quad Z_{f'} - \frac{1}{2} \frac{Z_f Z_{f'}}{\eta} = \pi_{n^*,f'}^{W^r}. \quad (15)$$

By subtracting the two previous equations, we obtain

$$Z_{f'} = Z_f + \pi_{n^*,f'}^{W^r} - \pi_{n^*,f}^{W^r}. \quad (16)$$

Without loss of generality, we order f and f' such that $\pi_{n^*,f}^{W^r} \geq \pi_{n^*,f'}^{W^r}$. Plugging (16) into (15), we obtain that Z_f satisfies the second degree polynomial equation (w.r.t. the variable Y):

$$Y^2 + Y \cdot [\pi_{n^*,f'}^{W^r} - \pi_{n^*,f}^{W^r} - 2\eta] + 2 \cdot \pi_{n^*,f}^{W^r} \cdot \eta = 0.$$

Note that this polynomial is strictly decreasing on the interval $[0, \eta]$ (we use here the normalization that $\pi_{n^*,f}^{W^r} \geq \pi_{n^*,f'}^{W^r}$). It implies that there is a unique suitable solution for Z_f and then from (16) also for $Z_{f'}$.

We have thus shown the induction step for any $n^* > 0$.

Case where $n^* = 0$. For $n^* = 0$, the expression of the probability that the player is sold to bidder f at the reserve price is slightly different (because winning at the reserve W_i^r does not imply that the valuation of the opponent is in the interval $[\Delta_0^{W_i^r}, \Delta_1^{W_i^r})$). We have

$$\pi_{0,f}^{W^r} = u_{0,f'}^{W^r} \cdot (1 - u_{0,f}^{W^r}) + (u_{1,f'}^{W^r} - u_{0,f'}^{W^r}) \cdot \left(1 - u_{1,f}^{W^r} + \frac{1}{2}(u_{1,f}^{W^r} - u_{0,f}^{W^r}) \right).$$

From the perspective of the previous argument, it is as if we replace $\pi_{n^*,f}^{W^r}$ by $\pi_{n^*,f}^{W^r} - u_{n^*,f'}^{W^r} \cdot (1 - u_{n^*,f}^{W^r})$ and we get then that there is a unique solution for $u_{1,f'}^{W^r}$ and $u_{1,f}^{W^r}$.

Remark: This part of the proof can be viewed as a “discrete version” with only two risks (i.e. two bidders in the auction interpretation) of the Pfaffian integral equations that appear in the generalized competing risk literature (see Meilijson (1981)). Without covariates, Athey and

Haile (2002) apply precisely Meilijson (1981) to get the identification of the English button auction without increments from the final winning price and the identity of the winner.

(2) We have shown that $H_f(\log[\Delta_n^{W^r}] - G_f(z_j, x_j))$ is identified for each f and $n \in \mathbb{N}$ for each covariates and reserve price on their support and once we have fixed the location $G_f(z_j, x_j)$. To get identification on \mathbb{R} of the CDFs H_f up to a location, it is sufficient to check that for the vector of covariates (z^*, x^*) , then any $\epsilon_f \in \mathbb{R}$ can be written as $\log[\Delta_n^{W^r}] - G_f(z_f^*, x_f^*)$ for some $W^r \in [0, \Delta(0)]$ and $n \in \mathbb{N}$.

First note that our regularity assumption guarantees that Δ_n^0 goes to infinity when n goes to infinity while $\Delta_0^0 = 0$. For any ϵ_f , there exists thus $n^* \in \mathbb{N}$ such that $e^{\epsilon_f + G_f(z_f^*, x_f^*)} \in [\Delta_{n^*}^0, \Delta_{n^*+1}^0]$. Since the function $x \rightarrow \Delta_{n^*}^x$ goes continuously (thanks to our regularity assumption) from $\Delta_{n^*}^0$ to $\Delta_{n^*+1}^0$ when the reserve price goes from 0 to $\Delta(0)$ (note that we use here that $\Delta_n^{\Delta(0)} = \Delta_{n+1}^0$). We obtain from the intermediate value theorem that there exists $w \in [0, \Delta(0)]$ such that $\epsilon_f + G_f(z_f^*, x_f^*) = \log(\Delta_{n^*}^w)$.

At this stage, we have identified the functions H_f once we have fixed the locations for the functions $G_f(z_f^*, x_f^*)$. Each location $G_f(z_f^*, x_f^*)$ is identified by the normalization $E[\epsilon_f] = 0$ for each f .

(3) As noted above, from the observation of the probability of sale in the auctions where only bidder f is “eligible”, we identify $H_f(\log[W_i^r] - G_f(z_{i,f}, x_{i,f}))$ for any possible point in the support of the vector of covariates $z_{i,f}, x_{i,f}$ and W_i^r . Since we have assumed that the support of the distribution H_f is \mathbb{R} , H_f is strictly increasing and we can thus identify G_f .

A.2 The probability $p_i^S(\mathcal{S})$

We express the probability $p_i^S(\mathcal{S})$ depending on whether the winning price is at or above the reserve price. If $w_i > W_i^r$ (which implies $|S| \geq 2$), we get from Bayesian updating:

$$p_i^S(\mathcal{S}) = \frac{\prod_{f \in S \setminus \{f^w\}} (\bar{H}_{i,f}^\Delta - \bar{H}_{i,f}) \cdot \prod_{f \notin S} \bar{H}_{i,f} \cdot \left((1 - \bar{H}_{i,f^w}^\Delta) + \frac{1}{|S|} \cdot (\bar{H}_{i,f^w}^\Delta - \bar{H}_{i,f^w}) \right)}{\sum_{\substack{S' \subseteq \{1, \dots, F\} \\ f_i^w \in S', |S'| \geq 2}} \prod_{f \in S' \setminus \{f_i^w\}} (\bar{H}_{i,f}^\Delta - \bar{H}_{i,f}) \cdot \prod_{f \notin S'} \bar{H}_{i,f} \cdot \left((1 - \bar{H}_{i,f^w}^\Delta) + \frac{1}{|S'|} \cdot (\bar{H}_{i,f^w}^\Delta - \bar{H}_{i,f^w}) \right)}. \quad (17)$$

The denominator corresponds to the probability (conditional on $x_{i,f}$ and $z_{i,f}$ for $f = 1, \dots, F$) that worker i is sold to f_i^w at price $w_i > W_i^r$.

If $w_i = W_i^r$, the set of active bidders at w_i possibly only contains $\{f_i^w\}$ (so here $|S| \geq 2$). The appropriate expression for the probability is as above except that in the denominator we also sum over sets that only include the winner. If $w_i = W^r$ we thus have for any $S \supseteq f_i^w$:

$$p_i^S(\mathcal{A}) = \frac{\prod_{f \in S \setminus \{f_i^w\}} (\bar{H}_{i,f}^\Delta - \bar{H}_{i,f}) \cdot \prod_{f \notin S} \bar{H}_{i,f} \cdot \left((1 - \bar{H}_{i,f_i^w}^\Delta) + \frac{1}{|S|} \cdot (\bar{H}_{i,f_i^w}^\Delta - \bar{H}_{i,f_i^w}) \right)}{\sum_{\substack{S' \subseteq \{1, \dots, F\} \\ f_i^w \in S', |S'| \geq 1}} \prod_{f \in S' \setminus \{f_i^w\}} (\bar{H}_{i,f}^\Delta - \bar{H}_{i,f}) \cdot \prod_{f \notin S'} \bar{H}_{i,f} \cdot \left((1 - \bar{H}_{i,f_i^w}^\Delta) + \frac{1}{|S'|} \cdot (\bar{H}_{i,f_i^w}^\Delta - \bar{H}_{i,f_i^w}) \right)}. \quad (18)$$

The denominator now corresponds to the conditional probability that worker i is sold to f_i^w at price $w_i = W_i^r$.

A.3 The auction rules: Accounting for RTM cards

In the IPL auctions of 2014 a large fraction of the players are RTM-eligible (see Table 2). It is therefore important to account for this phenomenon in our estimation procedure. RTM cards raise an issue concerning the definition of the pseudo-valuation of card holders. Indeed, the bidder who possesses a card to purchase player i may do so with or without using it. The number of cards being limited, the latter option may be preferable especially if the bidder has only a few cards left and wishes to save them to acquire other (highly valued) players later on in the auction. In principle we therefore have to distinguish two pseudo-valuations for this kind of bidder: one for purchasing i with the card, and one for purchasing this player without the card (the difference between the two would reflect the cost of using the RTM option). We cannot, however, estimate the pseudo-valuations of the second type. This is because in the data there is only one team which won an auction without using a card while it was actually RTM-eligible. Letting f_i^{elig} be the team eligible to use an RTM card, $V_{i,f_i^{elig}}$ thus corresponds to the pseudo-valuation of this team when player i is purchased via the card.

Next we outline how the likelihood function given in Section 2.6 should be slightly modified to account for RTM cards. We also give there the appropriate formula of the expected pseudo-valuation error term for team f_i^{elig} . We need to make two assumptions about the bidding behavior of f_i^{elig} , the team that is eligible to buy player i with an RTM card. First, f_i^{elig} does not participate in the bidding phase of the auction, and intervenes only (possibly, not necessarily) at the very end, after the auction process has reached the final price w_i . Second, f_i^{elig} matches the final price with its RTM card only if $V_{i,f_i^{elig}}$ exceeds w_i . The first assumption appears consistent with what we see in our data, namely that practically all RTM-eligible teams have won by using their cards (see Section 4). It implies in particular that f_i^{elig} does not belong to S , the set of bidders active at w_i . The second assumption states that, like any other bidder, f_i^{elig} only exercises the right to use the RTM card if it is willing to pay more for i than w_i .

Suppose that player i is a player who can be purchased with an RTM card. Let f_i^h be the team which wins the bidding phase of the auction for this player. Note that f_i^h necessarily belong to S but is different from f_i^{elig} (since, by assumption, the eligible bidder does not participate in the bidding phase of the auction). The bidder f_i^h has won the bidding phase either by winning the tie, or by being the unique bidder still active at $w_i + \Delta(w_i)$. The final winner of the auction, still denoted f_i^w , is f_i^{elig} if the eligible bidder uses its RTM card, and it is f_i^h otherwise. Let

$L_i^3(w_i, f_i^w, f_i^h; \alpha)$ denote the probability that player i is sold at price w_i to bidder f_i^w , and the bidder who wins the bidding phase of the auction is f_i^h . We have then:

$$L_i^3(w_i, f_i^w, f_i^h; \alpha) = \sum_{\substack{S \subseteq \{1, \dots, F\} \setminus \{f_i^{elig}\} \\ f_i^h \in S, |S| \geq 2}} \prod_{f \in S \setminus \{f_i^h, f_i^{elig}\}} (\bar{H}_{i,f}^\Delta(\alpha) - \bar{H}_{i,f}(\alpha)) \cdot \prod_{\substack{f \notin S \\ f \neq f_i^{elig}}} \bar{H}_{i,f}(\alpha) \cdot \left((1 - \bar{H}_{i,f_i^h}^\Delta(\alpha)) + \frac{1}{|S|} \cdot (\bar{H}_{i,f_i^h}^\Delta(\alpha) - \bar{H}_{i,f_i^h}(\alpha)) \right) \\ \times \left(\bar{H}_{i,f_i^{elig}}(\alpha) \cdot 1[f_i^w \neq f_i^{elig}] + (1 - \bar{H}_{i,f_i^{elig}}(\alpha)) \cdot 1[f_i^w = f_i^{elig}] \right).$$

The contributions of the type L_i^2 (contributions to the likelihood when i is sold at the reserve price) can be amended in an analogous way.⁴⁶ The control function term of the bidder that is eligible should be modified in the following way (omitting the match-specific indicator m):

$$CF_{i,f_i^{elig}}^\alpha = E[\epsilon_{i,f_i^{elig}} | \epsilon_{i,f_i^{elig}} < \log(w_i) - G_{i,f_i^{elig}}] \cdot 1[f_i^w \neq f_i^{elig}] + E[\epsilon_{i,f_i^{elig}} | \epsilon_{i,f_i^{elig}} \geq \log(w_i) - G_{i,f_i^{elig}}] \cdot 1[f_i^w = f_i^{elig}].$$

A.4 Construction of the performance measure

Our performance index is derived from the different player-specific performance measures observed during a cricket match. A cricket match involves two ‘innings’. An innings is a fixed-duration segment in the match, during which one team attempts to score (by *batting*) while the other team attempts to prevent the first from scoring (by *bowling*). The two teams have a single innings each, which is restricted to a maximum of 120 balls. The match proceeds with 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 at any given point of time is to post the maximum amount of score, called *runs*, in a certain period of time by striking the ball. A team’s innings is terminated when either 120 balls are bowled or when all the batsmen get dismissed by the bowling team. On each ball, the batsman can be dismissed (this is called a ‘*wicket*’), or score 0, 1, 2, 3, 4, or 6 runs. The team that posts the highest score wins the match.

We construct our performance index based on the following set of batsmen and bowler indicators. We use only those measures that are relevant to performance in a given match, rather than in the whole tournament⁴⁷. For robustness, we also consider a raw performance indicator, the batsman strike rate (defined below), as our dependent variable.

Batsman performance indicators: For all batsman indicators, higher values are associated

⁴⁶As mentioned in the main text, for one auction i the bidder eligible to use the RTM card won this auction without using it. The contribution to the likelihood is assumed to be $L_i^3(w_i, f_i^w; \alpha)$, given in Section 2.6, where f_i^w is the RTM eligible bidder and in $V_{i,f_i^{elig}}$ the dummy ‘Bidder eligible to use RTM’ equals one.

⁴⁷A commonly used performance measure in cricket is the ‘batting average’ which is constructed as the average runs scored by a batsman over several matches in a tournament. However we prefer using match-level data (and therefore consider only runs scored) to gain power and increase the precision of the estimates in our analysis. Admittedly this introduces more noise in the performance measure but we are able to net out some of this noise by extracting the common performance component from the match-specific performance indicators.

with higher performance.

- Batsman score (**BTS**): The total score of a batsman in a given match.
- Batsman strike rate (**BTR**): The average score of a batsman per 100 balls faced. This is equal to $[100 * (\text{Batsman score} / \# \text{ Balls faced})]$.

Bowler performance indicators: For all bowler indicators, lower values are associated with higher performance.

- Bowler economy rate (**BWE**): The average score conceded by a bowler per 6 balls. This is equal to $[\text{Bowler Score} / (\# \text{ Balls delivered} / 6)]$.
- # Wickets (**BWW**): The number of wickets taken by the bowler in the match. Note, that we take the negative of this value (with an addition of the constant 1) to preserve the interpretation of the index, i.e, lower values being associated with higher performance.

Based on these indicators, we construct two performance indices: a batsman performance index and a bowler performance index. To construct the batsman (bowler) performance index, we extract the common variation from the batsman (bowler) indicators using the method of principal component analysis and combine them into a single variable. For example the batsman performance index is derived by combining the standardized values of the indicators (denoted with a superscript S) where F_k is the weight the indicator variable K ⁴⁸:

$$BTP_i = F_{BTS} * BTS_i^S + F_{BTR} * BTR_i^S$$

Similarly the bowler performance index is given by:

$$BWP_i = - (F_{BWE} * BWE_i^S + F_{BWW} * BWW_i^S)$$

Since lower values for each indicator in the bowler index are associated with higher performance, we use the negative of the principal component score for bowler so as to make it comparable with the batsman performance index.

Quartile Measures of Performance: Finally, we assign a performance quartile to each player based on his batting or bowling performance index. The batting performance quartile is de-

⁴⁸More specifically, denote the $p \times p$ correlation matrix associated with the p batsman (bowler) performance indicators as C_{BT} (C_{BW}). The eigen (or spectral) decomposition of C_{BT} is $C_{BT} = F \tilde{\Lambda} F'$, where F is the matrix containing eigenvectors known as principal components while $\tilde{\Lambda}$ is a vector of eigenvalues (Jackson, 2003; Jolliffe 2002). Eigenvalues indicate the amount of variance explained by each component, whereas eigenvectors are the weights used to calculate components scores. In our analysis, we construct our index for both batsman and bowler performance using values/weights contained in the first eigenvector or principal component because it accounts for over 70% of the variation in the observed indicators. Therefore, our variable F_k represents the weight of the indicator K , based on the first eigenvector.

defined as the within match quartile rank⁴⁹ of a player based on the batting performance of all players (who have batted) in that match (with 4 representing the best performance and 1 representing the worst). We calculate a similar, quartile, measure for bowlers based on the bowler performance index.

We also combine the batsman and bowler indices to obtain a comprehensive measure of match performance for players who both bat and bowl. We do this by assigning the respective quartiles for batsman and bowlers, for specialist players. For all-rounders or players who are able to both bat and bowl, we assign the maximum quartile achieved between batting and bowling.

⁴⁹Our use of quartiles as the dependent variable, potentially, induces dependency across the set of players within each match. For this reason, we introduce match fixed effects to account for the induced match specific correlation.