Essays on Subjectivity in Hiring

by

Mark Azic

Bachelor of Arts, The College of New Jersey, 2012

Master of Arts, University of Pittsburgh, 2016

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Mark Azic

It was defended on

May 13, 2020

and approved by

Arie Beresteanu, Department of Economics, University of Pittsburgh

Federico Zincenko, Department of Economics, University of Pittsburgh

Rania Gihleb, Department of Economics, University of Pittsburgh

Kenneth M. Lehn, Department of Finance, University of Pittsburgh

Dissertation Director: Arie Beresteanu, Department of Economics, University of Pittsburgh

Essays on Subjectivity in Hiring

Mark Azic, PhD

University of Pittsburgh, 2020

This dissertation consists of three essays on the role that subjective evaluation plays in hiring markets. The first chapter looks at the part subjectivity plays in determining which workers match with which firms. The second chapter looks at how providing hiring managers with subjective information about a job applicant changes managers' evaluations. Finally, the third chapter devises a method to determine the level of disagreement between managers in how they interpret a subjective piece of information.

Table of Contents

List of Tables

List of Figures

Preface

My dissertation was inspired by two of the best parts of my grad years: the friends I made in the econ department, and the Pitt students that I taught. Seeing friends who were of equal academic quality enter the job market, only to experience wildly disparate outcomes, put the idea in my head that much of what happens in hiring markets is dictated by a randomness that's unobservable and difficult to measure. But it wasn't until I was teaching an undergrad econometrics course, in which some very bright students asked about how worker-firm sorting impacted subjective labor market data, that I got the idea for how to come up with a measure for the impact of that randomness – which I labeled, "subjectivity." Many thanks to my friends and students; I'll miss them very much.

I was lucky to have a dissertation advisor who encouraged my research, even if it wasn't a perfect match for his own. Although he's never explicit about it, Arie has a certain philosophy about how economics should be done. It stresses thinking about assumptions at every stage of work, but also not shying away from asking creative and original questions. I don't know if it's a *reflection* of me choosing him as my advisor or of his impact on me, but I realize now that my own ideas of how economics should be done are in line with his. Much appreciation to Arie for his guidance on all matters economics and beyond.

Finally, I want to thank my family; nothing means more than when they say they're proud of me.

1.0 Introduction

This dissertation consists of three chapters, each of which, in its own way, looks at how subjectivity impacts hiring. I define subjectivity to be the part of assessment in hiring decisions for which no objective criteria is present, and explain why it is potentially important to several labor market phenomena, including inequality and the rising importance of "soft" skills.

Oyer and Schaefer (2011) comment on the lack of empirical studies on hiring, calling the field a "black box". Despite the lack of studies, almost every sorting model is forced to make strong assumptions on the hiring behavior of firms. This is understandable, as the focus of most models is on the choices of workers. If assumptions on hiring behavior are erroneous, however, then so are the inferences drawn from these models. By focusing on the important role that subjective information plays in hiring markets, this dissertation contributes to economists' understanding of how hiring markets work.

In Chapter 2, I look at the importance subjectivity plays in labor market hiring, focusing on the role it plays in determining who gets hired, and by whom. To measure the extent and importance of subjectivity, I construct a two-sided matching model of the labor market based on [39], and structurally estimate it using techniques from the industrial organization literature [23]. I apply the model to the labor market for professional baseball player contracts, a labor market which avoids many challenges to identification and which also provides me with unmatched agents (i.e. unsigned players). I show how these unmatched agents can be used to identify and estimate the distribution of subjectivity, which I model as an unobservable. I find that subjectivity assessment in hiring is pivotal to 5% of the market, in that their employment or unemployment status hinges on subjective evaluation.

While Chapter 2 looks at how subjective assessment plays a role in determining *who* gets hired, it says nothing about how the presence of subjective information changes manager evaluations. This is done in Chapter 3. Chapter 3 uses laboratory experiments to simulate a labor market and implement exogenous variation in the types of information available to employers. It shows that subjective information increases the overall valuation managers assign to worker profiles, but it doesn't improve the managers accuracy in either worker valuation or selection. Furthermore the study finds that subjective information strongly reduces the well-known bias in favor of male workers when it comes to hiring for a stereotypically male tasks. This implies that subjective information is a potentially useful tool in combating discrimination, but it doesn't provide any efficiency gains.

Chapter 4 concludes the dissertation by helping to understand some results in Chapter 3, and by providing a methodology for determining how subjective a covariate is. In Chapter 3, I recover the distribution for a covariate when it is subjective, i.e. the value that it takes for the decision maker is idiosyncratic in a way that is unobservable to the researcher. To accomplish this, I make use of three things. The first two are the experimental treatment setting and the availability of an unbiased proxy for the subjective covariate, both of which are provided the setting in Chapter 3. The third is a set of results from the measurement error literature which allow me to recover the distribution of subjectivity via a deconvolution argument. I look at the distribution of subjectivity in a hiring setting, where hiring managers receive written statements from workers. I find that the quality of these written statements causes great disagreement amongst hiring managers; for the median quality statement, a 25/75 ratio for how managers interpret the statement quality shows that managers can disagree by 30%. With firms collecting a large amount of subjective information on job applicants (e.g. interviews), my result implies that subjectivity is an important consideration for worker-firm matching studies.

2.0 Subjectivity in Hiring: Theory with Evidence from Professional Baseball Contracts

2.1 Introduction

Labor markets are rife with imperfections. Over the past fifty years, economists have made strides in building ever-realistic models which capture many of these. However, virtually no models of worker-firm sorting allow for the fact that many hiring decisions depend – at least in part – on subjective evaluation. For example, there is usually no objective criteria to judge the answer to a job interview question by (e.g. "tell me a little about yourself"). Therefore, it is completely plausible – and anecdotally common – for two identical hiring agents to reach differing conclusions about an applicant.

This paper measures these differing conclusions, and the extent to which they drive worker-firm sorting. To do so, I focus on a specific hiring market: professional baseball player contracting. This labor market provides all firms with the same objective measures of applicant quality, allowing me to overcome identification issues caused by asymmetric information. Combining theory from the matching literature with recent tools from empirical industrial organization, I build a structural model of worker-firm sorting. Estimates of the model show that subjectivity is pivotal for 5% of the labor market, in the sense that for 5% of workers, whether they are employed or unemployed depends on the sort of subjective assessment they receive.

Previous research has demonstrated the importance of hiring on long-run labor market outcomes.[1] As such, the part subjectivity plays in hiring decisions can speak to several labor market phenomena. For example, subjectivity can explain why observationally equivalent workers receive such different labor market outcomes — a question which underlies the literature on residual wage inequality (e.g. [36]). Another field where hiring subjectivity is significant is the literature on the growing importance of social and "soft" skills on labor outcomes (e.g. [18]); it's likely that important traits such as work ethic, sociability, and leadership are judged subjectively during the hiring process. Finally, subjectivity is important to understanding the role "luck" plays in the labor market.[2] In many hiring instances luck can take the form of a favorable (or unfavorable) subjective review.

Despite this relevance, this paper – as far as I know – is the first to look at the effects subjectivity has on worker-firm sorting.[3] Studying hiring presents challenges, especially with respect to controlling for inter-firm differences in information and objectives. Not only is it rare for firms to be seeking the exact same qualities in an applicant, but each firm also employs unique screening and interview processes, leading to different information collected between firms. Problems remain when looking within a firm, as different managers involved in a hiring decision might view an applicant at different times, have different levels of input on the hiring decision, or have competing motivations. Finally, there are major complications on the applicant side, as applicants self-select into which jobs they apply for.

[1]For an incomplete list see [64], on the importance of job-switching on wages; [1] for the first of many studies using matched employee-employer data to show the importance of firm effects; and [37], [54], and [3] showing that graduating from college during a recession has negative and persistent effects on earnings, as graduates place in lower quality jobs initially.

[2]For a recent discussion of the role luck plays in labor markets, see [42] and the Volume 36, S1 special issue of the *Journal of Labor Economics*, titled, "Firms and the Distribution of Income: The Roles of Productivity and Luck".

[3]In the Literature Review I contrast this paper with ones which allow worker-firm match quality to vary across firms, and models in which firms learn a worker's match quality.

I overcome these obstacles by focusing on a specific hiring market: contracting between professional baseball teams and players. In Major League Baseball (MLB), baseball teams and players contract with one another in a process called Free Agency, where each player is free to negotiate a contract with any one of the 30 teams. The only players eligible for Free Agency are those with at least six years of professional experience, and a player's performance during those six years is public. Therefore, this setting overcomes information challenges, as all employers form expectations on potential hires using the same information set. In addition to limiting information asymmetries, Free Agency also minimizes search frictions; each team has complete understanding of which players it can match with, and communication between the two sides is inexpensive. Finally, professional baseball is a setting where employers have a rich set of objective measures for each potential hire. I'll argue that the presence of these strong objective measures makes my estimates for subjectivity a lower-bound for several important labor markets (e.g. school teachers and corporate executives).

While the specifics of Free Agency help to identify subjectivity, the setting does pose challenges. One challenge is that baseball teams contract with multiple players in each period, and the substitutability between those players is unclear. For example, players are differentiated by the position they play. Assuming all else equal (e.g. salary, ability, health), two players who play different positions may complement one another better than players who play the same position. Any model approximating Free Agency will need to allow teams to have flexible preferences over sets of players. Another challenge comes from what is unobservable in Free Agency. Although I see which players and teams match and a contract's terms (i.e. salary), rejected offers are unobserved; I do not see that a team was originally interested in another player or the salary the team was willing to pay for that player. Therefore, any model used to measure subjectivity needs to be able to do so using only data on observed matches.[4]

In response to these challenges, I estimate the importance subjectivity plays in contracting by turning to the matching literature. The basis for my model is [39], a two-sided matching model with money transfers, in which firms (or in my case, baseball teams) have preferences over the *set* of workers (baseball players) hired. Although Kelso and Crawford's model has been highly influential in the theoretical matching literature, there has been no previous attempt to structurally estimate it (so far as I know). By doing so, I expand the empirical discrete matching literature to labor markets in which firms hire multiple workers (discussed further in Related Literature).

My model modifies [39] by allowing firms to have an additive unobservable – representing subjectivity – in their preferences over sets of workers. The model is solved using the same deferred-acceptance algorithm as Kelso and Crawford: over discrete rounds, teams bid on players simultaneously; players with multiple bids reject offers, and command higher bids from the teams they rejected; finally, the auction concludes when no offers are rejected.

The solution concept for the model is that of core stability. Core stability requires that (i) all agents are entering matches which are preferable to not matching at all (i.e. their outside option); and that (ii) no coalition of teams and players can break off from the observed matches, and negotiate prices which leave them all better off. After making an assumption which restricts the complementarity between players, I show that the auction mechanism produces a matching which is core stable.

I then apply core stability to the Free Agency market. Each year, Free Agency features players who go unsigned (i.e. unmatched agents). That each team had

[4] This is similar to a second-price auction where only winning bids are observed.

the option of signing these players – but chose not to – gives a series of preference relations on the surplus a team could expect had it added each unsigned player. Together, these preference relations allow me to establish nonparametric identification for the distribution of subjectivity. Having identified the distribution of subjectivity, I then estimate the model in two steps.

In the first step, I parameterize the deterministic component of firm surplus. Concerned about unobserved heterogeneity, I modify the matching maximum score estimator (MSE) of [23] and [24] to allow for a finite mixture of types. This modification allows me to provide conservative estimates for subjectivity. Using the modified matching MSE, I estimate the portion of team (i.e. firm) preferences based on observables. Having estimates for the deterministic component of team match-surplus, I then estimate the distribution of subjectivity in the second step. The approach in the second step closely follows the identification argument: the estimated parameters from the first step allow me to estimate the deterministic component of match surplus between each team and each unmatched player; a revealed preference argument (i.e. that the team decided not to match with the player) yields an inequality for the subjectivity draw underlying that potential match. Combining these inequalities across the entire set of unmatched players gives a likelihood.

Related Literature

Labor economics has developed several modeling frameworks to explain wage differences among ex-ante identical workers. Macro-focused models have introduced dynamic frictions to produce wage differences. For example, in the job matching model of [35] worker-firm match-quality is an experience attribute; firms only learn match-quality over time, and two ex-ante identical workers can have different match-quality, producing divergent outcomes. In the search literature models of on-the-job search create wage asymmetries as a result of the stochastic timing of outside job offers (e.g. [13]). While models like [35] and [13] are useful in modeling and explaining phenomena regarding the aggregate labor market, neither allows for subjectivity. Instead, both frameworks focus on competitive market assumptions, which subjective hiring would violate[5].

While perfectly competitive models of worker-firm sorting produce results that are elegant and useful in explaining many labor market phenomena, they're also inconsistent with the data on wage differences, with [41] empirically testing and rejecting competitive theories of wage determination. Instead of adopting a perfectly competitive framework, I use the finite matching framework previously employed to study problems such as marriage and medical residency matching[6]. The decision to base my model off of [39] allows me to model worker-firm sorting as the solution to a bargaining game. I'm able to show that this bargaining game can accommodate subjectivity in a tractable manner, and therefore that this approach to modeling (small) labor markets is attractive to researchers.

Methodologically, the papers most similar to this one are [24] and [25]. [24] uses a matching model to structurally estimate the deterministic component of bidder valuations in FCC spectrum auctions. FCC spectrum auctions were conducted via a simultaneous ascending auction, which makes analysis via traditional structural auction methods computationally challenging. The authors recognize that the simultaneous ascending auction represents a generalization of a matching model. Using this insight, they adapt the matching maximum score estimator of [23] to estimate the deterministic component of bidder valuations — an estimation procedure that

[5] This is also true of empirical models. For example, sophisticated search models such as that in [58] allow for heterogeneous employers and employees, stochastic match-quality shocks, and on-the job search. However, such models implicitly assume that when two identical employers see the same applicant, there is no disagreement over the applicant's quality. For more recent examples in empirical search where this assumption is made, see [46], and [45]

[6] See [60] for a summary.

mirrors step one of my estimation procedure. However, unlike [24], my primary object of interest is the distribution of the unobservable component of match quality. I show how the unobservable distribution can be obtained by imposing additional model structure regarding the simultaneous ascending auction's form, and by taking advantage of having unmatched agents.

[25] study the identification of unobservables in matching, using only information on who matches with whom. They provide strong intuition into the identification challenge, and give results showing the usefulness of unmatched agents in identifying unobservables. [25] rely on having many markets for implementation. By contrast, I use an approach that is easier to implement, but which requires stronger assumptions; namely, I assume the matching mechanism takes the form of Kelso and Crawford's algorithm.

With respect to the content of my application, I believe this paper's explicit focus on subjectivity in hiring to be novel, with the exception being noted in the following paragraph. Subjectivity has been previously studied in the literature on compensation[7] where it has been shown to compress earnings between workers. Subjectivity has also been used to explain trading in markets that rely on public information (e.g. the stock market)[8]; while famous "no-trade" results (e.g. [4] and [51]) argue there should be no trading in markets with common information, subjective interpretation of that information offers an explanation for trading.

While these applications of subjectivity are interesting, they're not particularly informative about the role subjectivity plays in hiring. Seeking to remedy this, my coauthor and I study the question of subjectivity in hiring in [6], using an experimental labor market setting. Altering the subjective information a hiring manager

[7] For examples, see [52], [11], [12], and [7]

[8] See [32], [66] and [38]

is given on a candidate, we find that subjective information increases the overall valuation managers assign to candidates, but that it does not improve manager accuracy in either worker valuation or selection. Furthermore, we find that subjective information strongly reduces the well-known bias in favor of male workers when it comes to hiring for a stereotypically male task.

Outline

The rest of this paper is structured as such. Section 2 introduces the professional baseball contracting setting and provides data describing the contracting process. After introducing the application setting, I go through the model in Section 3, describing components of the model in terms of the baseball application (e.g. players and teams are used in light of workers and firms). After setting up the model, I apply the algorithm of Kelso and Crawford to obtain a solution. I conclude Section 3 by showing that the distribution of subjectivity is nonparametrically identified using unmatched agents. In Section 4 I describe my two-stage estimation procedure. Section 5 provides estimates for the deterministic and subjective components of match surplus, along with counterfactual simulations regarding assignment and salary determination. Section 6 concludes and discusses the relevance my study of subjectivity has to other labor markets.

2.2 Background and Data

I now describe the key features of Major League Baseball free agency, emphasizing those which help identify subjectivity. I then introduce the dataset I constructed and provide statistics describing the market.

2.2.1 Setting

Major League Baseball (MLB) players are contracted employees who play for the team that owns their contract. Players without an active contract are labeled *free agents*, and Free Agency (FA) is the market through which free agents are able to sign a new contract with any one of the thirty MLB teams.

The terms of FA are set via collective bargaining. MLB players are represented by the MLB Players Association, a union which negotiates a single collective bargaining agreement (CBA) that covers player-team relations for all MLB teams.[9] Each year, FA begins on a set date,[10] with the first five days of FA representing the "Quiet Period". During the Quiet Period, teams and free agents are free to discuss the merits of contracting with one another, however negotiating the terms of a contract (e.g. salary) are explicitly prohibited, and doing so can result in sanction under the the CBA. Therefore, the Quiet Period can be thought of as a period of discovery, wherein the two sides learn about each other.

Following the Quiet Period, teams and players explicitly negotiate with one another. In conversations with agents who represent players, this has been described as similar to bidding, with agents obtaining an offer from one team, only to request that other suitors raise their offers. While players do care about non-monetary factors (e.g. location, coaching, teammates), salary is the dominant feature in any contract. This is consistent with the short earnings career of most players, as the average player plays only 5.6 seasons (see [67]), and earnings during the first six seasons of a player's career are kept artificially low (described in more detail shortly).

[9]The period I study, 2006-2016, covers three CBAs: 2003-2006, 2007-2011, and 2012-2016. Across the three CBAs, very little with respect to FA changed. The complete CBAs can be viewed at the Players Association's website: http://www.mlbplayers.com/ .

[10]FA begins at 9 a.m. Eastern Time on the day following the conclusion of the World Series (i.e. the day after the season concludes).

As a labor market, FA allows for efficient search. In many labor markets, firms and workers may only have partial knowledge of who they can possibly match with.[11] By contrast, in FA all teams and players are completely aware of each other. Players are represented by professional agents,[12] making communication between the two sides inexpensive. Furthermore, while many labor markets are characterized by job vacancies and unmatched workers appearing on a rolling basis, FA has a set starting time; I make use of this in my model. Finally, during the 2006-2016 period, FA was generally seen as a well-function market, with robust bidding.[13] Together, these features will provide a measure of subjectivity that is less tainted by search frictions than it would be had I chose to study another labor market.

The sort of contracts negotiated in FA are remarkably similar.[14] All contracts must pay at least a minimum salary, the level of which is stipulated by the CBA. For the period I study, this minimum ranges from \$380k to \$535k, increasing an average of 4.7% per year, and never decreasing. While contracts are not explicitly guaranteed, voiding a contract is extremely difficult;[15] notably, any injuries a player sustains while playing for a team do not change the money he is owed. Some contracts do include annual incentives, and options for future years. Incentives are usually tied to the number of games a player appears in, and therefore offer a hedge against injury or especially poor performance. Options for future years can belong to the team,

[11]For example, a firm looking to hire a custodian in New York City will not see all workers who may qualify for and be interested in the job.

[12]Player agents must pass a licensing test and screening administered by the Players Association. This is done to assure that agents have a strong understanding of the CBA, and to endow agents with fiduciary responsibility over the player they represent.

[13]I don't include data past 2016, as recent years have been characterized by far weaker bidding, with the Player's Association accusing teams of colluding with one another to keep player salaries low.

[14]The CBA has a contract template which many contracts follow closely (CBA 2012, Appendix A).

[15]See Chapter 1 of [56] for examples of how extreme circumstances must be for a contract to be voided.

player, or both sides. I summarize the contract data in the next subsection.

That contracts are effectively guaranteed makes it possible to interpret salaries as a function of the expected contribution a player will provide to their team (as there is little possibility to renegotiate salary later on). To determine a player's expected contribution, teams rely on baseball's rich set of objective performance measures. Performance in baseball is hyper-measured, with every throw, swing, and catch digitized.[16] Each baseball game consists of a finite series of states, with each state resulting in a discrete number of outcomes. This structure makes keeping statistics especially easy, as does the solitary nature of performance (e.g. at any given time in a game, only one player is throwing the ball, and only one player is attempting to hit it.).[17] In addition to there being many objective measures of a free agent's performance, there are also many years of it: the only players eligible for FA are those with at least six years of professional experience.[18]

It is important to understand who is using all of these measures to determine a free agent's value. Teams employ dozens of analysts charged with evaluating players. While many of these analysts have a background in baseball (e.g. as former players or coaches), teams also employ analysts with backgrounds in economics, statistics, computer science, and business – workers who are capable of sophisticated data

[16]For an example of just how digitized performance is, consider *PITCHf/x*, which was installed in every stadium beginning in 2007. *PITCHf/x* is a series of cameras which record twenty pieces of data for every pitch thrown, including velocity, three dimensional movement, and the angle of the pitcher's arm throughout the throw.

[17]Baseball's obsession with measurement has sprouted the creation of multimillion dollar data companies devoted to statistics (e.g. STATS LLC), as well as membership organizations: (e.g. SABR: Society for American Baseball Research). [62] provides a history of the sport's obsession with statistics, describing the attraction baseball's numeracy holds for many fans.

[18]Prior to that, the CBA makes it extremely difficult for a player to become a free agent. Players are essentially tethered to their initial team for the first three years of their career, during which they are paid a salary near the league minimum. For years 4-6 of a player's career, salary is determined through an arbitration process, which is seen as heavily favoring the player's team retain his services, and at a below-market rate.

analysis.[19] From the perspective of what a team spends on players, coaches, stadium rights, etc., the salaries paid to analysts are minuscule ([57]). Therefore, differences in analyst capacity across teams are immaterial. Finally – and most importantly – all teams have access to the same player performance data. Most all statistics are available for free online. Those which are not available online can be purchased, at a cost that is far less the operating budget of any team. Therefore, the information set that each team conditions on when forming projections about a player's quality can be thought to be roughly the same. That teams have symmetric information on potential hires is a far cry from most other labor markets, where heterogeneous screening procedures lead to information asymmetries. By eliminating such asymmetries, my measure of subjectivity is better identified.

2.2.2 Data

I collect information on all players who filed for FA from 2006 through 2016. Data is taken from Baseball-Reference, Baseball Prospectus, and Spotrac. In total, I observe 1436 guaranteed contracts[20] signed (i.e. matches between players and teams) and 320 players who go unsigned (unmatched agents). Each year represents a "market", with roughly 160 players available for teams to bid on.

For each player, I observe their previous team and their new team; the length of the contract they sign; and the salary guaranteed each year. I also observe detailed production measures for each past season the player played. For each team I ob-

[19] During the early 2000s teams became increasingly dedicated to using advanced data techniques. This was covered heavily in the press, most notably in the book *Moneyball* ([43].

[20] I consider a contract to be guaranteed if any money is guaranteed. Teams will also sign some players to *minor league contracts* which generally contain no guaranteed money. I do not focus on these contracts, as the players who sign them are usually ones with little to no experience playing, and therefore unknown commodities to most teams.

serve performance from the previous season (e.g. record of wins and losses, amount scored); team spending on player contracts; and measures of team talent at each position. While I do not observe team revenue directly, I do observe team attendance, which tracks revenue closely and correlates with differences in team spending power (see [28]).

Although my data contains an abundance of performance measures – both for players and for teams – it's impractical to work with so many measures. Therefore, going forward, I focus on just one index meant to capture performance: *Wins Above Replacement* (WAR). The idea behind WAR is to summarize a player's entire contribution toward helping his team win. A player's WAR value is the amount of wins that player contributes to his team's success *above* what the team would have achieved had the player been replaced by an available (i.e. unsigned) player. Baseball involves discrete events, and WAR is calculated by measuring how a player's performance during each of those discrete events affects his team's probability of winning. While there is some disagreement over how to calculate WAR, it closely tracks more traditional measures of performance (e.g. home runs) and is widely studied and used.[21] Table 1, below, characterizes players and teams on the market.

[21]For reference, a WAR of 1 roughly corresponds to an average MLB player; the best handful of players will post a WAR of 6+. For more details, see [28].

Table 1: Characteristics of Market Participants

Panel A: Players

	Mean	**25th%**	**Median**	**75th%**	**SD**
Guar Sal (millions, 2006 dollars)	14.42	1.53	4.12	13.18	29.00
Age	33.07	31	33	35	3.433
Talent (in WAR)	0.943	0.15	0.68	1.42	1.2
Years	9.45	7	9	11	3.46
Re-Signed (proportion)	0.231				
1 year contract (proportion)	0.601				
2 year contract	0.214				
3+ year contract	0.185				
Number of signed players	1436				

Panel B: Teams

	Mean	**25th%**	**Median**	**75th%**	**SD**
Attendance (in millions)	2.54	2.00	2.48	3.05	0.69
Team talent (in total WAR)	32.72	24.7	32.1	41	10.77
Amnt spent (millions, 2006 dollars)	62.75	15.34	37.01	82.61	74.78
Number of players signed	4.35	3	4	5	2.19
Number of team-year observations	330				

Panel A of Table 1 focuses on the 1436 signed players in my data. The data on salary – which is the discounted sum of guaranteed dollars circa the year 2006 – shows great variation and that salaries are heavily right skewed. Notably, salaries are skewed more so than player talent, implying a non-linear relationship between the two; small talent differences among players lead to large differences in pay. Also worth noting in Panel A are the descriptors for Age and Years played. As mentioned earlier, FA is only an option for players who have played at least six years professionally. The players in my sample are ones for whom each team has a long history of past performance to consult.

At the bottom of Panel A are some characteristics of contracts. The first of these shows that players are much more likely to re-sign with their current team (23%) than to switch to any one of the other 29 teams. While this could be a sign of players preferring to stay in place – or of a player's current team having better information than outsiders – it could also be an outcome endogenous to the marketplace; if a player becomes a free agent, then he leaves a "hole" in his former team's roster, as his position is now empty. Therefore, his old team will seek out someone to fill that spot. I will account for this in my model by allowing the value a team places on a player to be a function of the team's composition. The bottom of Panel A also makes it clear that the majority of player contracts are short term, with over 83% lasting one or two years. This provides support for the assumption that players enter the market with uncertain future earnings, and therefore that it is optimal for players to maximize guaranteed income, even in the face of nonmonetary preferences.

Panel B characterizes the teams in my data, with each observation at the team-year level (e.g. Baltimore-2006). On average, teams sign just over four players to guaranteed contracts in a given year, with those contracts guaranteeing $63 million to players. Like player salaries, the total spending of a team is heavily right-skewed, much more so than team attendance or talent (measured as a sum of WAR across players the team has under contract) are.

I now use the data to demonstrate the potential importance of subjectivity in this market. The ideal argument for the presence of subjectivity would be seeing the contract each team is willing to offer to each player, and then arguing that discrepancies are the result of subjectivity. Unfortunately, I only see data on signed contracts, and therefore only have a lower bound for what *one* team is willing to pay for a given player. Therefore, what I instead turn to is explaining player salaries using observable data. Table 2 does this, using a series of regressions where a player's salary is the dependent variable.

The four regressions in Table 2 show that much of a player's salary cannot be explained via observables. Controlling for salary differences across positions in each regression, Column 1 shows a simple linear regression of salary on a player's characteristics leaves much unexplained. Column 2 adds quadratic terms, and explains more of the variation in salary. Note that the large change in the coefficient on WAR corresponds to what the descriptive statistics showed: at the high-end, small differences in talent lead to large differences in pay. Column 3 adds controls for the team a player matches with; perhaps heterogeneity across teams explains much of pay. Finally, Column 4 includes year fixed effects, in case certain years in the data are irregular.

What the regressions show is that the portion of salary explained by observables is limited. Attempts at further regressions, using higher degree polynomials or more traditional statistics of player performance than WAR (e.g. home runs) does little to change this. What else could be explaining pay which hasn't been captured in the above regressions?

There are three determinants of pay which are not captured in the above regressions and which I would argue are important to determining pay. The first is complementarity between a player and the set of other players a team is signing. If a team has signed a talented set of players, perhaps they are willing to overpay to obtain another talented player, and assure their earlier expenditures do in fact result in a strong team. A second likely determinant of player pay which is not captured in the above regressions is substitutability between a player and the other players on the market. If many players of the same position and of similar talent are available in the same market, competition should drive pay lower for each of them. Unlike hedonic models, which assume a dense market, FA matches a discrete number of players and teams; competition (or the lack of it) will cause differences in pay.

Finally, the third potential cause for the differences in pay is also the topic of this paper: subjectivity. I now introduce a model which incorporates all three of these aspects.

2.3 Model

Motivated by the setting of baseball Free Agency contracting, I construct a two-sided matching model based on [39]. The model treats each year in my data as a separate two-sided matching market. Markets are independent of one another, and each market is characterized by the model below.

2.3.1 Market Participants and Preferences

On one side of the market are teams, which are denoted by $j = 1, ..., N$, and on the other side are players, denoted by $i = 1, ..., M$. The two-sided market is many-to-one, with each team able to match with multiple players in a period but each player able to match with at most one team; both sides also have the option of going unmatched.

I now describe the preferences of each side, beginning with an assumption on player preferences.

Assumption 1: *(i) Players have preference over salary only. (ii) All players have a reservation salary below the league minimum salary,* s_{min}.

Assumption 1 allows me to avoid having to determine whether a player and a team match with one another due to (i) the player having a strong liking for the team; (ii) the team having a strong liking for the player; (iii) both player and the team liking one another; or (iv) the player or team disliking everyone else they can match with. Assumption 1 is also realistic for most players. As detailed in the previous section, the overwhelming majority of players sign one-year contracts, experience short careers, and see their income vary greatly from year to year. Therefore, assuming players maximize guaranteed income fits well for most of the market.[22]

Teams have preferences over sets of players, salaries paid to players, and a subjective shock. Let C^j denote the set of players signed by team j; $\mathbf{s}_j$ denote the vector of salaries, with each component $s_{i,j}$ denoting the salary team j pays player i; and $\mathbf{L}_j$ denote a subjective valuation shock. Team j's total surplus from signing the set of players C^j given salaries $\mathbf{s}_j$ and subjective shock $\mathbf{L}_j$ is given by the function $u(C^j, \mathbf{s}_j, \mathbf{L}_j)$.[23] I make the following assumptions about the subjective shock compo-

[22]Assuming that players maximize salary — rather than a more complicated surplus function — also makes the marketplace resemble an auction, as players are matched with their highest bidder. The next subsection introduces this mechanism.

[23]Note that a team's preferences will also be a function of that team's characteristics (e.g. players

nent to team surplus:

Assumption 2: *(i) The team surplus function is additively separable in L:* $u(C, \boldsymbol{s}, L)$ *equals* $= u(C^j, \boldsymbol{s}_j) + \boldsymbol{L}_j$. *(ii) Team j's subjective shock* $\boldsymbol{L}_j$ *can be decomposed as* $\boldsymbol{L}_j = \sum_{i \in C^j} \mu_{i,j}$*, where* $\mu_{i,j}$ *represents the subjective valuation shock team j receives for player i. (iii) Each* $\mu_{i,j}$ *is an iid draw across both teams and players, with each* $\mu_{i,j}$ *being drawn from the continuous distribution* F_μ*.*

Assumption 2 plays the important role of describing the shape subjectivity is allowed to take. Together, conditions (i) and (ii) make it so that a team's subjective evaluation of a group of players is entirely based on the team's subjective evaluation for each individual player in that group. This restricts subjectivity. For example, the complementarity between the players in C^j is not something that can be evaluated subjectively; instead complementarity between players signed must be entirely captured through $u(C^j)$.[24]

Condition (iii) says that each team has a separate subjective evaluation for each player in the market, but those evaluations are all drawn from a common distribution. This iid assumption is strong. In the estimation section it will become apparent that some parts of the iid assumption are testable (e.g. high quality and low quality players drawing identical shocks), while others are not (e.g. shocks not being correlated within a team). Recovering F_μ will be the main goal of estimation.

I now describe the auction setting used to match teams and players.

currently on the roster). To simplify notation, I suppress this for now, before making it explicit when discussing estimation.

[24]Requiring the stochastic component of surplus to be separable is a common assumption in both empirical models of two-sided matching and multi-unit auctions (e.g. [16] and [29], respectively).

2.3.2 Auction Mechanism

Players are allocated to teams following a simultaneous ascending auction mechanism. Prior to the auction beginning, each team j receives subjective shocks $\mu_{i,j}$ for all players on the market, $i = 1, ...M$.[25] Each team then uses these shocks, along with the deterministic $u(C, \mathbf{s})$, to compute its surplus for all possible $(C, \mathbf{s})$ pairings. These computed surpluses then form the basis for bidding behavior in the auction.

The rules of the auction are as such. The auction proceeds as a sequence of discrete rounds of bidding. In each round t, a team sees a price vector, $\mathbf{s}(t)$, which specifies the bids that team needs to make to become the leader for any one of the N players. The team then chooses $M^j\big(\mathbf{s}(t)\big)$, where $M^j\big(\mathbf{s}(t)\big)$ is the solution to the team's problem:

$$M^j\big(\mathbf{s}(t)\big) = \underset{C}{\operatorname{argmax}}\{u\big(C, \mathbf{s}(t)\big)\}.$$

In words, $M^j\big(\mathbf{s}(t)\big)$ gives team j's most preferred set of players when the current asking prices are given by $\mathbf{s}(t)$.

In the first round of the auction, $t = 0$, $s_{i,j}(0) = s_{min}$ for all players. Teams then choose $M^j\big(\mathbf{s}(0)\big)$. Each player who receives multiple offers rejects all but one, with ties broken randomly. If team j made an offer to player i in round t and had it rejected, then in round $t+1$, j's price for i goes up by one dollar: $s_{i,j}(t+1) = s_{i,j}(t)+1$.[26]

Rounds continue in this fashion until a round in which there are no rejected offers. At that point, the auction concludes, with players allocated to the team that has the highest bid for them and that bid representing their salary.

[25] Whether a team's subjective shocks, $\mu_{i,j}$, are private information or not is immaterial, as I assume straightforward bidding in what follows, with teams not conditioning on their rivals behavior. I discuss this at the end of the section.

[26] The important part here is that bids rise by some discrete amount; whether it be by \$1 or some other amount is immaterial.

This simultaneous auction mechanism is nearly identical to the one described in [39], and closely resembles the deferred-acceptance algorithms of [26] and [27]. Solving a simultaneous ascending auction generally requires assumptions which restrict the complementary between goods (see[50] for an overview). I now introduce one such assumption, Gross Substitutes:

Assumption 3 (Gross Substitutes): *Let* $\boldsymbol{s}$ *and* $\bar{\boldsymbol{s}}$ *be two salary vectors such that* $\boldsymbol{s} \leq \bar{\boldsymbol{s}}$*. Define* $T^j(C^j) = \{i \in M^j(\boldsymbol{s}) | \bar{s}_i = s_i\}$*. Then* $T^j(C^j) \subset M^j(\bar{\boldsymbol{s}})$*.*

In words, Gross Substitutes says start out with a set of prices $\mathbf{s}$ and choose your favorite set of players, $M^j(\mathbf{s})$. Now raise the prices of some players, while keeping the price for others the same. Gross substitutes requires that you still want the players in $M^j(\mathbf{s})$ whose price stayed the same.

With respect to baseball's Free Agency setting, Gross Substitutes is similar to the assumption that when you bid on a player, you submit a contract offer to him, and you cannot withdraw that offer. If the player receives a better offer from another team, he will reject your offer, at which point you can decide whether you want to make another offer for the player, at his new, higher price. But so long as his price remains the same (i.e. no one else bids on him), you're required to honor your salary offer to the player.

Mechanically, Gross Substitutes acts to further limit complementarity between players. For example, a team might want a particular pitcher and catcher because together they're especially valuable (e.g. they have terrific teamwork), so the team bids on both players. Tomorrow, if the price for the pitcher becomes too expensive, but the catcher's price remains the same, the team needs to still desire the catcher. Gross Substitutes prevents the market from unraveling in instances like these, as it tethers teams to a player so long as their bid is highest for that player.

Although restrictive, nearly all multi-unit auctions of heterogeneous goods require some assumption limiting substitution. In addition, this assumption doesn't seem too restrictive in the marketplace of baseball Free Agency. In conversation with several sports agents who represent players in Free Agency, they said it was almost unheard of for a team to make an offer to a player and then to withdraw it.[27]

Also worth noting is that while Gross Substitutes restricts complementarity, the restrictions are on the $u(C, \mathbf{s})$ function. Therefore, Gross Substitutes does not restrict subjectivity; the only assumptions on subjectivity remain those outlined in Assumption 2.

2.3.3 Solving the Model

I now define the terms used to describe the model's solution.

Definition 1: An *allocation* consists of an assignment rule, which matches players to teams, and a salary vector, which specifies what each player is paid.

Definition 2: An allocation is *individually rational* if it (i) ensures that no player is accepting a salary offer that is less than his reservation wage; and (ii) no team is making negative surplus, given the set of players they sign, C^j, and the salary vector $\mathbf{s}$.

Definition 3: A *core allocation* is an individually rational allocation such that there exists no coalition of players and teams, and a salary vector $\mathbf{r}$, such that: (i) $r_{i,j} \geq s_{i,j}$

[27]One reason for this is that it would send a poor signal to the agent representing the player, as well as to other agents, many of whom work for the same sports agency firm as the offended agent. Therefore, the threat of receiving a "black-mark" punishment and harming future business dealings makes contract withdrawals rare in this setting.

for all i in the coalition; and (ii) $u(C, \mathbf{r}) \geq u(C, \mathbf{s})$ for all teams in the coalition.

These definitions are standard in the matching literature. Definition 1 defines an allocation, which can be thought of as an auction outcome. Definition 2 refers to players and teams in the marketplace having the outside option of matching with no one (i.e. going unsigned and signing no one, respectively); an allocation being individually rational assures that no agent is doing something less preferred than their outside option. Definition 3 gives the solution concept of the core, which assures that there does *not* exist a coalition of players and teams that can break off and negotiate amongst themselves in a way which leaves (i) all players in the coalition weakly better off and (ii) all teams in the coalition weakly better off.

I now get to the main two theoretical results regarding the solution of the model.

Theorem 1 (Kelso and Crawford): *Under Assumptions 1-3, the auction mechanism results in a core allocation after a discrete number of rounds.*

Proof: See Appendix A.1.

The proof of Theorem 1 closely follows Theorem 1 of [39]. While Theorem 1 proves the existence of a core allocation, it does not prove its uniqueness. I do this, by adding the following assumption:

Assumption 4 (Unique Tie-Breaker): For each player i, there exists a strict preference ordering, $\succ_i$, such that for any two salary offers where $s_{i,j} = s_{i,j'}$ either $s_{i,j} \succ_i s_{i,j'}$ or $s_{i,j'} \succ_i s_{i,j}$.

Up until this point, it was assumed that players had preferences over salaries only, and that a player receiving identical salary offers breaks the tie randomly. Assumption 4 gets rid of random tie-breaking, by assuming that each player – conditional on receiving the same salary offer from two different teams – has a strict preference over the two offers. Note that players still have preferences over salary first and foremost. Instead of altering this, Assumption 4 is akin to giving players lexicographical preferences, in which the first dimension is salary, and the second dimension is something arbitrary (e.g. geographic distance from current location).[28]

The benefit of Assumption 4 is that now, players who receive multiple offers of equal value, will reject the same offer each time. This means that for every set of subjective shock realizations, there will be a unique core allocation.

Theorem 2 (Uniqueness): *Under Assumptions 1-4, the auction mechanism results in a unique core allocation after a discrete number of rounds.*

Proof: See Appendix A.2.

I now use the individual rationality requirement implied by Theorem 1 to prove identification of F_μ, the distribution of subjectivity. The usefulness of Theorem 2 will be made clear in the second stage of estimation.

2.3.4 Identification of F_μ

The distribution of subjectivity, F_μ, is identified by combining the individual rationality of teams (which is implied by Theorem 1) with the presence of unsigned free agents. It also requires two further assumptions: one on functional form, and

[28]Note that what this second dimension is can be heterogeneous across agents, i.e. each player i can use a different tie-breaker. All that Assumption 4 says is that each player does have a deterministic tie-breaker.

the other an exclusion restriction.

Assumption 5: *The $u(C^j, \boldsymbol{s})$ component of team surplus can be expressed as the sum of a Ψ term and $\big(v_j(i) - s_{i,j}\big)$ where Ψ_j gives the synergy between members of C^j and $v_j(i)$ gives the standalone value of each player i in C^j.*

Given Assumption 5, consider an unsigned player, u and for now assume that $v(.)$ is identified and estimated outside of the model. Theorem 1 shows that the auction will result in a core allocation, which means that the allocation is also individually rational. Therefore, if a team, j, decided not to sign player u – even at the minimum salary – then it must be that:

$$[\Psi_j(C^j) - \Psi_j(C^j \cup u)] \geq [v_j(u) + \mu_{u,j} - s_{min}], \quad (1)$$

i.e. that player u's addition to C^j harms synergy (LHS) more than his standalone surplus can justify (RHS). I introduce the following exclusion restriction.

Assumption 6 (Exclusion Restriction): *Each unsigned player u has two observable components, X and Z, allowing the player to be denoted as $u(X, Z)$. Observable component Z affects $v(.)$ but not $\Psi(.)$, so that for $Z \neq Z'$: $\Psi\big(C \cup u(X, Z)\big) = \Psi\big(C \cup u(X, Z')\big)$, and $v\big(u(X, Z)\big) \neq v\big(u(X, Z')\big)$.*

The identification argument follows from this restriction. Consider a set of unsigned players who differ in Z but not in X: $\{u(X, Z_1), u(X, Z_2), ..., u(X, Z_N)\}$. That team j decided against signing these players implies that for each one:

$$[\Psi(C^j) - \Psi(C^j \cup u(X, Z))] - v(u(X, Z)) + s_{min} \geq \mu_{i,j}. \quad (2)$$

By the exclusion restriction, variation in Z – but not X – changes the LHS of (2) only via changes to $v(.)$, which was assumed to be estimated. Hence, the CDF of $\mu_{i,j}$ is pinned down via variation in Z among unsigned players.

2.3.5 Discussion

In this section I discuss some of the assumptions and limitations of my model, and relate them to auction modeling.

Simultaneous ascending auctions (SAA) were first used by the Federal Communications Commission to auction off segments of radio spectrum.[29] The idea behind using SAAs was that the FCC had little idea how different pieces of spectrum should be packaged; bidders may regard certain segments as complements to one another, whereas other segments were substitutes for one another. At the same time, structuring auctions as ascending allows for price discovery, reducing bidder anxiety from "the winner's curse".

Although SAAs are widely used, their theoretical properties are not well understood. Much of this has to do with the difficulty synergy between different items introduces to solving for optimal bidding strategies. As such, obtaining theoretical results usually requires an assumption limiting complementarity between auction items, and/or limiting bidding strategies.

My model makes two such assumptions. The first assumption, Gross Substitutes, limits the shape synergy can take, by tying a bidder to an object, so long as his most recent bid for that object is the highest bid standing. This is in fact realistic for most

[29]The FCC's use of a SAA was seen as successful, and SAAs have been used ever since by governments around the world to auction off resources, raising hundreds of billions of dollars in revenue; see [49] for a non-technical summary, and [50] for a more technical treatment.

simultaneous ascending auctions: placing a bid on an object obligates you to purchase that item so long as your bid is highest (i.e. its price does not rise). However, the problem it introduces is with respect to what is called the "exposure problem". The exposure problem is that a bidder who sees two items as complements runs the risk of winning an object she no longer wants once she is outbid for the other item. The Gross Substitutes assumption restricts a team's surplus function in a way so as to eliminate the exposure problem, and therefore puts a strong restriction on auction behavior.

The second strong assumption I make on bidder behavior in a SAA is that of straightforward bidding. Bidders in my model do not strategically attempt to forecast the bids of other teams when placing their own bids. Instead, a team looks at the price of each player it can bid on at the start of each round, compares it to what surplus that player would generate for the team — conditional on who else the team currently has outstanding bids on — and decides whether to bid or not. This is much more restrictive than canonical auction papers, in which bidders condition on rivals' characteristics when forming strategies. The reasons for this are realism and practicality.

Conditioning on rivals when forming a bidding strategy is considerably more difficult in my setting than in the typical auction model. The reason for this is the scale and multi-unit setup of a baseball team's predicament. An average free agent market contains 130 players, and 30 teams. That means, a baseball team must not only form its own valuations over 2^{130} sets of players[30], but also must do so for each of it's 29 rivals, and then condition on that information in each round of bidding. Therefore, straightforward bidding is far more realistic and practical for both the

[30] That's 1.36×10^{39} sets of players.

baseball executive submitting bids and for the economist studying those bids.[31]

This paper contributes to a new empirical literature on multi-unit auctions. [29] develop and estimate a structural model of bidding in simultaneous first-price auctions, where bidders have preferences over combinations. They impose a similar additive separability on bidder surplus to the one I use, not allowing for complementarity to interact with a bidder's idiosyncratic, standalone value for a good. They model a framework in which all bids are observed, and then use symmetric bidding strategies and first-price auction theory – along with an exclusion restriction similar to my own – to identify and estimate the model, in the style of [31].

[40] addresses complementarity across heterogeneous items, but does so in a setting of sequential auctions, in which the result of the first auction affects bidding in a second auction. Her topic of interest is the level of complementarity between parcels of land. By making use of the sequential format and observing all bids, she's able to condition on bids in the first auction to identify and estimate the value of synergy, allowing her to make less restrictive assumptions on the bidder-surplus function.

Finally, [24] take a similar approach to me, in that they relate a simultaneous ascending auction to a matching model. Like my model, [24] assume additive separability between the standalone, complementary, and stochastic components of bidder surplus. Unlike my model, they put fewer restrictions on bidding behavior and on the mechanism through which the simultaneous ascending auction takes place. This

[31]Of course, in the actual market, teams do take some considerations over the players other teams will bid for, and the probability that they will be able to win one auction they participate in does affect how aggressively they bid in other auctions. That teams do, in practice, make such considerations, should limit the harm the exposure problem poses to my model. However, at the same time, not factoring in strategy is likely to introduce some bias into estimates. Note that the incremental, straightforward bidding approach also rules out jump-bidding, in which bidders can signal to others the intensity of their preferences. While jump-bidding certainly occurs for the premier handful of free agents, conversations with agents representing players at the lower end of the market suggest that incremental bids are more common for the majority of players.

comes at a cost, as they're not able to prove a core allocation or its uniqueness – though the different application and goals of their paper make these theoretical results less important.

2.4 Estimation

Estimation follows a two-step procedure, in which the first step estimates the deterministic component of team surplus and the second step estimates the distribution of subjectivity, F_μ. Both steps rely on a revealed preference approach.

For each FA market, I assume the researcher observes who matches with whom, and at what salaries (i.e the final *allocation* of the market); the researcher does *not* observe any offers made by teams to players, except for the offers which the two sides contract on. In addition, the researcher observes a set of covariates describing each player and team in the market. Let x_i represent a vector of covariates for player i and y_j represent a vector of covariates describing team j. The matrix X_j collects the vector y_j with the set of vectors $\{x_i | i \in C^j\}$.[32]

2.4.1 Step 1: Estimating Deterministic Surplus via MSE

Recall that by Assumptions 2 and 5, team surplus can be expressed as $\Psi(X_j) + \sum_{i \in C^j} \big(v(x_i, y_j) + \mu_{i,j} - s_{i,j} \big)$. Define $\bar{u}(X_j)$ as the deterministic component of team surplus, net of salary: $\bar{u}(X_j) \equiv \Psi(X_j) + \sum_{i \in C^j} v(x_i, y_j)$. I assume that $\bar{u}(X_j)$ is

[32]Markets across years are independent of one another. Therefore, in my notation I drop any time subscript.

known up to a finite vector of coefficients, β, which I write as $\bar{u}_\beta(X_j)$. Step 1 uses the semiparametric estimator developed in Fox (2018) to estimate β, while placing no parametric restriction on F_μ.

[23] adapts the maximum score estimator (MSE) of [48] to matching games with transfers. The estimator relies on a rank order condition, in which the difference in probabilities between two possible matchings occurring depends on their difference in expected utility. [24] point out that the rank order condition can be seen as the econometric analog to the matching solution concept of pairwise stability. Pairwise stability requires that for any two pairs of matches, the agents involved prefer the matching they are in to the counterfactual match implied by a swapping of partners; in this sense, the current matchings are stable when considering any two pairs.

By Theorem 1, my model of FA results in a core allocation, in which no coalition of any size can find an alternative allocation which leaves all of its members better off. Therefore, my model implies pairwise stability, and I assume the rank order condition holds in my estimation.[33]

Unlike [23] and [24] who focus on estimation when match transfer data is absent, I observe the salaries players and teams contract on. To make use of this transfer data – the benefits of which will be clear in the results – I adopt the matching MSE approach used in [2].

The MSE objective function is constructed via the following revealed preference argument. Consider two teams, $\{j, j'\}$ and two players, $\{i, i'\}$, such that $i \in C^j$ and

[33][24] do not prove pairwise stability, as their model is much more general than mine. However, they do provide evidence from both theoretical and experimental work to argue that pairwise stability is likely to hold in simultaneous ascending auctions. This evidence also applies to the use of pairwise stability in my setting.

$i' \in C^{j'}$. Because team j chose to match with player i at salary $s_{i,j}$, it must be that:[34]

$$\bar{u}_\beta(X_j) - s_{i,j} \geq \bar{u}_\beta\big(X_j \cup \{x_{i'}\} \setminus \{x_i\}\big) - s_{i',j} \tag{2.1}$$

where $\bar{u}\big(X_j \cup \{x_{i'}\} \setminus \{x_i\}\big)$ represents the deterministic surplus team j would enjoy, were it to swap out player i with player i'; and $s_{i',j}$ denotes a counterfactual salary that j would need to pay player i' to attract him.[35] What this counterfactual salary would need to be is unobserved in the data. However, by Assumption 1, players have preference over salary alone. Therefore $s_{i',j} \geq s_{i',j'}$. Furthermore, a team's surplus is strictly decreasing in salary paid. Together with the discrete auction format, this implies that $s_{i',j'}$ is equal to the highest salary – within a dollar – a team other than j' was willing to pay for player i' (similar to paying the second highest price in an English auction). That player i' received $s_{i',j'}$ in equilibrium, and not a dollar more, implies that $s_{i',j}$ in inequality (1) could be replaced by $s_{i',j'}$ – give or take a dollar.[36] Therefore, although I do not observe $s_{i',j}$ in the data, in what follows, I substitute $s_{i',j'} = s_{i',j}$ into inequality (1).

In addition, the revealed preference argument above implies that for team j' to match with player i' instead of i, it must be that:

$$\bar{u}_\beta(X_{j'}) - s_{i',j'} \geq \bar{u}_\beta\big(X_{j'} \cup \{x_i\} \setminus \{x_{i'}\}\big) - s_{i,j'} \tag{2.2}$$

where the unobserved counterfactual salary $s_{i,j'}$ can be substituted by the observed

[34] This ignores subjectivity – the stochastic component – for the time being.

[35] Note that the salaries of other players in C^j remain fixed during the hypothetical swap, and therefore cancel out in (1).

[36] The model assumed that salary bids went up by a dollar each round of bidding in which a player received multiple bids. This was done for convenience, as bids rising by any discrete amount would have solved the model in the exact same way. In practice, of course, actual bids don't rise by a dollar during each round of bidding. However, it is common to see salaries at the low-end differ by just \$10k. Therefore, it may be safer to replace $s_{i',j}$ with $s_{i',j} + \$10k$. However, even this difference is too small to matter in estimation. Therefore, I use the simplification, $s_{i',j} = s_{i',j'}$.

allocation salary $s_{i,j}$. Together, inequalities (1) and (2) give the following objective (score) function:

$$Q(\beta) = \sum_{j=1}^{29} \sum_{j'=j+1}^{30} \sum_{i=1}^{|C^j|} \sum_{i'=1}^{|C^{j'}|} \mathbf{1}\{\bar{u}_\beta(x_j, y_{C^j}) - \bar{u}_\beta(x_j, y_{C^j/\{i\}\cup\{i'\}}) \geq s_{i,j} - s_{i',j'}\}$$
$$\times \mathbf{1}\{\bar{u}_\beta(x_{j'}, y_{C^{j'}}) - \bar{u}_\beta(x_{j'}, y_{C^{j'}/\{i'\}\cup\{i\}}) \geq s_{i',j'} - s_{i,j}\}. \quad (2.3)$$

Given a parameter vector β, the score function in equation (3) counts how often inequalities (1) and (2) are *both* satisfied in the data, considering all such possible swaps. The leftmost summation fixes one of the baseball teams, j. The next summation fixes another one of the 30 teams, j', for which player swaps with j will be considered. The third leftmost summation chooses a player i on team j, while the the last summation compares i to a player on team j'. The estimate, $\hat{\beta}_{MSE}$, is the parameter vector which satisfies the most inequality pairs. Because the score function only considers the deterministic components of surplus, not all inequalities will be satisfied, as the unobserved subjectivity terms also play a part in driving observed matches.

Although the score function in (3) is written so as to consider *all* possible swaps, doing so is impractical and unnecessary; [21] shows that consistent estimates of β_{MSE} can be obtained by considering a large and random subset of all possible choices (i.e. swaps).

2.4.2 Identification of β

For the identification of a team's deterministic surplus function I use a parameterization.[37] I make the following, linear functional form assumption on $\bar{u}_\beta(X_j)$:

$$\bar{u}_\beta(X_j) = \beta_C X_j + \sum_{i=1}^{|C^j|} \big(\beta_A x_i y_j + \beta_S x_i\big), \tag{2.4}$$

where the first component, $\beta_C X_j$, captures the complementarity between the players in C^j with one another (e.g. positional overlap amongst players signed); where the second component, $\beta_A x_i y_j$, captures the assortative matching between each individual player in C^j and the team j (e.g. more talented players contribute more to higher revenue teams); and where the last component, $\beta_S x_i$, captures the standalone contribution of each player, which is valued identically across teams.

In transferable utility matching models where transfer data is absent, the two inequalities in score function (3) are often replaced by just one inequality:

$$\bar{u}_\beta(X_j) + \bar{u}_\beta(X_{j'}) \geq \bar{u}_\beta\big(X_j \cup \{x_{i'}\} \setminus \{x_i\}\big)\bar{u}_\beta\big(X_{j'} \cup \{x_i\} \setminus \{x_{i'}\}\big). \tag{2.5}$$

This inequality – which is implied by pairwise stability – says that total surplus must be higher in the realized matches than under a swap. By using the inequality in (5) instead of the pair of inequalities in (3), the score function would not be able to identify the β_S term, and is, therefore, unidentified. Therefore, the data on salaries is important in that it allows me to identify the standalone term, the significance of which becomes clear in the Results section.

Another benefit of observing salary data is that it provides a natural scale normalization. The coefficient on $s_{i,j}$ is set to -1 in the team surplus function, giving

[37][22] studies nonparametric identification of the matching surplus function in many-to-many games, using pairwise stability to derive results.

other β coefficients an interpretation in terms of dollars.

Finally, two other key identification assumptions, both introduced in [48], are existence of a *super-regressor* and *median independence*. The super-regressor assumption requires that there be at least one covariate which, conditional on all other covariates, has a continuous support equal to the real line. Median independence requires that conditional on any covariate, the unobservable component of surplus have a median of zero. [21] extends the median independence assumption from Manski's binary choice problem to the multinomial setting, resulting in the rank order condition assumed earlier in this section, which requires that matches with higher deterministic surplus have a greater chance of occuring.

In my setting, the super-regressor assumption is required to achieve point identification of β; the existence of a continuous regressor acts to break ties in deterministic surplus between different β values, translating to a unique β_{MSE}. I employ several continuous variables in my application. While using a finite amount of data means that no regressor in my dataset satisfies the super-regressor assumption in the strict sense, repeated estimates yield unique β coefficients giving the maximal score.

The rank order assumption puts fairly innocuous restrictions on the shape of subjectivity. Assumption 2 assumed that subjective draws μ_i were *iid* from a continuous distribution F_μ. These are sufficient conditions to satisfy the rank order assumption. Furthermore, the rank order condition also allows for heteroskedasticity across baseball teams; the distribution of F_μ can be different across team characteristics, so long as each draw from a conditional subjectivity distribution $F_{\mu|j}$ is *iid*. However, some distributions of subjectivity do not meet the rank order condition. One such case is heteroskedasticity across player characteristics. For example; if the subjective distribution for players who are Pitchers has a larger right tail than the subjective distribution for Catchers, then a comparison of a Pitcher and a Catcher with identical deterministic surplus would lead to the Pitcher being chosen more often – in violation of rank order.[38]

2.4.3 Step 2: Estimating F_μ Using Unmatched Agents

The estimation approach here is constructive, following the identification result laid out in Section 3.4.

In Step 1, $\bar{u}_\beta$ was consistently estimated. In this step, those consistent estimates are used along with unsigned players to obtain F_μ. I return to the notation of the model, in which team surplus can be written as $\Psi(X_j)+\sum_{i\in C^j}\big(v(x_i,y_j)+\mu_{i,j}-s_{i,j}\big)$.

Let $u=1,...,U$ be the set of unsigned players. The fact that team j did not want to add player u and pay him the minimum salary implies that:

$$\begin{aligned}\Psi(X_j)+\sum_{l\in C^j}&(v(x_l,y_j)+\mu_{l,j}-s_{l,j})\\&\geq \Psi(X_j\cup\{u\})+\sum_{l\in C^j}(v(x_l,y_j)+\mu_{l,j}-s_{l,j})+(v(x_u,y_j)+\mu_{u,j}-s_{min}),\end{aligned}\quad (2.6)$$

which can be simplified to:

$$\Psi(X_j)-\Psi(X_j\cup\{u\})-v(x_u,y_j)+s_{min}\geq \mu_{u,j}. \quad (2.7)$$

Define $\bar{\Delta}(j,u)$ to be equal $\Psi(X_j)-\Psi(X_j\cup\{u\})-v(x_u,y_j)+s_{min}$, the left hand side of equation (5); i.e. the deterministic cost to team j of not signing player u at s_{min}. Player u was not signed by team j, and by every other team. The probability of this is given by:

[38] [21] discusses the rank order assumption in detail, and shows why models of random coefficients would also fail to meet the assumption.

$$\Pr(u \text{ is unsigned}) = \Pi_{j=1}^{30} \Pr(\mu_{u,j} \leq \bar{\Delta}(j,u))$$

$$= \Pi_{j=1}^{30} F_\mu\big(\bar{\Delta}(j,u)\big). \tag{2.8}$$

This argument can be extended across each unsigned player, giving:

$$\Pr(u{=}\mathit{1},\ldots,U \text{ is unsigned}) = \Pi_{u=1}^{|U|} \Pi_{j=1}^{30} F_\mu\big(\bar{\Delta}(j,u)\big). \tag{2.9}$$

Equation (9) provides a likelihood function which can be used to estimate F_μ. By Theorem 2 of the model, a given set of subjective draws for all pairs $\{i,j\}$ in the market, produces a unique allocation outcome. This implies that the likelihood in (9) is well-defined.[39] Although non-parametric estimation is possible, data limitations make it infeasible. As such, my results assume that F_μ takes the parametric form of an exponential distribution.

2.4.4 Discussion

Assumption 1 simplified the economic model by assuming that players cared *only* about salary. This assumption was then used in Section 4.1, where it allowed me to replace a swap's unobserved counterfactual salary with the signed player's actual salary plus a dollar; if player i only cares about salary and signs with team j for $s_{i,j}$ then another team j' must not have valued player i at that price, else they would have bid up his salary. Here I discuss how robust my estimates are to players having

[39] In simultaneous ascending auctions, both existence and uniqueness of equilibrium is difficult to prove, with the exception of special cases, leading to assumptions on both being necessary for estimation. My model – although stringent – requires me to make no such assumptions in my estimation.

preference over more than just salary, first considering what impact this may have on estimates of β and then what impact it may have on estimates of F_μ.

Let $r_{i,j}$ represent player i's non-pecuniary distaste for playing with team j. This consists of preferences for things like weather, differing endorsement opportunities, city specific amenities, etc. and may vary across both teams and players. Consider the inequality in (1) which considers the swap of player i for i' from the perspective of team j. The argument below inequality (1) says that although the hypothetical salary $s_{i',j}$ is not observed, the fact that players care only for salary makes it so that the hypothetical $s_{i',j}$ can be replaced by $s_{i',j'}$; player i' only cares about salary, so he could be gotten by j for just a dollar more.

Allowing for players to have preferences over more than just salary – allowing $r_{i,j}$ to differ across teams and players – means that this is no longer the case. Now, the hypothetical salary needed to lure i' in the swap must satisfy:

$$s_{i',j} \geq s_{i,j'} + [r_{i',j} - r_{i',j}].$$

The inequality above reflects that the counterfactual salary to be paid would need to be the salary that i' actually received from j' plus a compensating differential based on the difference between his distaste for playing for teams j and j'; if i' prefers team j in non-pecuniary matters, then j can sign him more cheaply, and vice versa.

Define $\delta_r^{i'}(j, j') = r_{i',j} - r_{i',j}$. Assumption 1 of my model sets $\delta_r^{i'}(j, j') = 0$ for all possible swaps, while my estimation approach ignores $\delta_r^{i'}(j, j')$, relegating it to the unobservable in Step 1 of estimation. This is concerning if ignoring $\delta_r^{i'}(j, j')$ leads to an unobservable in Step 1 which violates the rank order assumption needed for consistent estimates of β. However, this is unlikely to be the case. Consider two randomly drawn pairs, (i, j) and (i', j') and their hypothetical swap. By construction, it is just as likely that $\delta_r^{i'}(j, j')$ takes on a positive value as it is to take on a negative value. Therefore, $\delta_r^{i'}(j, j')$ should have a median of zero, and satisfy the median independence.

Allowing players to have preferences over non-pecuniary items is also unlikely to hamper estimation of F_μ. Recall that Step 2 is based only on the set of players who go unsigned. Then $r_{i,j}$ represents player i's reservation salary for playing in city j. The second part of Assumption 1 assumed that all players have a reservation salary below s_{min}, which is likely true for most players, for whom earnings are precarious and the time frame to earn in is extremely short. In this case, there will be no bias in estimates of F_μ.

In the case that some players do have $r_{i,j}$ which lies below s_{min}, then these players are incorrectly being included in the likelihood function given by equation (9); equation (9) assumes that the team received a subjective evaluation for these players which lead to the team deciding not to offer s_{min} when in reality it was the player's decision to remain unsigned due to his high reservation wage. This introduces a bias which likely shifts the distribution of F_μ to the left, as some of the unsigned players may have received favorable subjective draws which made teams interested in them, only to have a team's offer declined. Again, it is unlikely that this is true for many of the unsigned players in my sample, and I expect this bias to be minimal.

Likelihood approaches are the dominant estimation technique for parametric auction models. However, it would be difficult for me to use a likelihood to estimate my entire model. While teams have preferences over sets of players, I only observe data on the winning bid for each team. Unlike the classic, single-good, second price auction (i.e. bid your value), bidding behavior is not easily characterized in the simultaneous setting. Constructing a likelihood on preferences over sets of players, using only the data on winning bids, creates an infeasible combinatorial problem. Furthermore, constructing a likelihood over how likely a given allocation of players and salaries is, is also a problem, as I observe data on just 11 matching markets.

Another common approach to estimation, the first order approach (e.g. [31]), relies on being able to write-out optimal bidding strategies in terms of the distribution of valuations. Bidding data can then be mapped non-parametrically to valuations. This is the approach taken in the simultaneous first-price auction in [29]. However, this approach will not work for my model, as it requires a more strategic environment as well as complete data on bidding, which I do not have.

The MSE approach that [23] takes allows me to overcome the two challenges of estimating deterministic surplus: its scale and its lack of complete data; while the presence of unmatched agents allows me to go beyond the existing matching literature, and obtain estimates for the unobservable component of surplus.

2.5 Results

2.5.1 Estimates of β

I now estimate $\bar{u}_\beta(X_j)$, based off the parameterization in equation (4). Note that in that equation, the coefficient on salary is assumed to be -1. This assumption acts as a normalization, while also giving the coefficients on β an easy to interpret form.

To capture assortative matching between player i and team j, I define variables *WarAtt* and *TalOver*. *WarAtt* is equal to the player's talent – measured in Wins Above Replacement (WAR) – multiplied by a team's attendance in the previous year, in millions. This variable accounts for the tendency of more talented players to match with higher-revenue teams, where their talents may be worth more. The variable *TalOver* measures the difference between a player's talent and the talent of the player on team j who plays the same position. A player may represent more of an "upgrade" on some teams than other. *TalOver* allows me to capture that. For *TalOver* and all other variables regarding position, I group players into five positional groups based on where they played most in the most recent season (Catcher, Infielder, Outfielder, Starting Pitcher, Relief Pitcher).[40] Looking at the parameterization in equation (4), the coefficients on *WarAtt* and *TalOver* both fall within β_A.

To capture the synergy between the set of players a team signs, C^j, I define variables *PosComp* and *TalComp*. The idea behind both variables is to capture the composition within set C^j. *PosComp* measures the degree of positional overlap within C^j. There is no perfect way to do this, as an "ideal" composition of signings' positions does not exist, and will vary by team. The proxy I use for an ideal positional composition signed is based on the signings of teams with the highest attendance; these are the teams with the highest revenue, and therefore most likely to sign a set of players that represents the ideal positional composition.[41] For each year, I compute the average number of signings at each position for the five highest attendance teams in the market. I define these as the "ideals". I then take the sum of the square differences between the number of players at each position in C^j and their respective ideal.[42] The square term captures that over – or under – signing a position by a wide margin is not commonly observed in the data (e.g. no team signs 4 catchers in one market). The variable *TalComp* is defined to be the absolute

[40] This grouping was to account for the fact that many infielders and outfielders will switch from one position within the infield or outfield, to another (e.g. Second Base to Third Base).

[41] The results for *PosComp* are nearly identical if the highest attendance teams are replaced with the highest spending teams.

[42] Note that this captures the fact that certain positions get signed more frequently than others (e.g. starting and relief pitchers get signed more frequently than catchers) as this is accounted for in the measure of "ideals".

difference between a player's talent and the talent of the players in C^j, sans that player. This captures teams signing sets of players who are of similar talent level. The coefficients on *TalComp* and *PosComp* comprise the β_C term in equation (4).

Finally, the last set of variables I define are those that comprise the standalone terms in equation (4), β_S. These are variables which describe a player i but do not interact with the team, j, or any of players signed by the team, C^j; their surplus is common across teams. Here I include covariates for player talent, age, years of experience, and a dummy for whether a player re-signed with the team which previously had him under contract.

Table 3 provides estimates for the coefficients in β_A and β_C. Estimates are based on the score function in equation (3). Rather than consider all possible pairs of player-team swaps, I follow Fox (2007), who shows that not all swaps need be considered, so long as a large enough set it. I randomly draw 5000 pairs of player-team swaps, with each pair coming from the same market, and involving two different teams. The score function was maximized using a differential evolution algorithm. For each specification, the algorithm was run 15 times, with Table 3 reporting the estimates which maximized the score.

The table shows four specifications, with the first only including the variables meant to capture assortativity and complementarity. The coefficients provided are in units of \$100 thousand, multiplied by the unit of the variable. This makes interpretation extremely untidy.[43] As the focus of this paper is subjectivity and not baseball itself, I do not discuss any coefficients in detail.[44]

[43]For example the coefficient on *WarAtt* in specification (1) says that every one increase of *player-wins-above-replacement-millions-of-team-attendees*, makes a team's match surplus go up by \$42 thousand.

[44]Baseball aficionados or anyone else curious about the meaning of my results are welcome to contact me; I'd be happy to discuss the baseball meaning of my results further.

The last row of Table 3 shows the Score, which is the percentage of times *both* inequality pairs in the score function given by equation (3) are satisfied. This provides a measure of "fit" for the model.[45] Unsurprisingly, as more covariates are included in the model, the fit improves. The inclusion of standalone variables is shown to significantly increase the fit, illustrating the importance of their being identified. The reduced form evidence in Table 2 showed the importance of quadratic terms; as such, the specification in column (4) adds squared values for each covariate, bringing the Score to 0.468. However, in subsequent specifications, including additional variables had virtually no effect on the Score. While it did increase, the increases were generally 1 to 2% per additional covariate. To avoid over-fitting, I take the specification in column (4) as my final estimate of equation (3). Step 2 of estimation, which draws conclusions on subjectivity, will be based on this specification.

2.5.2 What Does β Say About Subjectivity?

Note that with estimates for β in hand, I can calculate the deterministic component of match surplus for each unsigned player, as well as for each signed player. For unsigned players, this is given by the equation for $\bar{\Delta}(j, u)$ defined in Step 2 of the Estimation section.

In a Free Agency market where subjectivity was completely turned off (i.e. $\mu_{i,j} = 0 \ \forall i, j$), an unsigned player would go signed whenever $\bar{\Delta}(j, u)$ was greater than zero. I calculate $\bar{\Delta}(j, u)$ for all 301 unsigned players in my dataset, and find that for 38 of them there is at least one team for which $\bar{\Delta}(j, u)$ is positive, i.e. that would have signed the player. Therefore, for 12.7% of unsigned players, a purely de-

[45]Note that it would be wrong to interpret it analogously to R-Squared.

terministic world would have meant being signed. Repeating this exercise for all 1436 signed players is similarly interesting. It shows that for 43 signed players (3.0%), no team has a positive deterministic match surplus for them, i.e. they would have gone unsigned sans subjectivity. In total, these results imply that for roughly 5% of the entire market, subjectivity was pivotal to their either being signed or going unsigned.

2.6 Conclusion

In this paper I present a unique approach to a new question: how much does subjective assessment affect labor market sorting. I adopt the [39] model from the theoretical matching literature and show how it can be structurally implemented using techniques from the industrial organization literature. As such, I provide a new modelling approach for researchers who are interested in understanding the preferences of the employer side of the labor market, which has traditionally been neglected in empirical models of labor-firm sorting.

Estimating firm hiring preferences and the part subjectivity plays in them is a task riddled with identification challenges, as most labor markets are characterized by firms with extremely idiosyncratic hiring objectives and screening processes, as well as complications stemming from the applicant side, such as search frictions and self-selection into where to apply. To avoid these identification problems, I look at the very narrow labor market setting of professional baseball player contracting. The institutional features of this labor market provide for better identification with respect to the aforementioned challenges to studying hiring.

The baseball labor market also provides me with unmatched agents, who I show are important in the identification and estimation of subjectivity. After estimating the model and the distribution of subjectivity, I show that for roughly 5% of the overall market, subjectivity is pivotal, in that it's existence explains their employment or unemployment. The baseball labor market I study is unlike many other labor markets economists study. However, its rich set of objective measures of worker performance and the quantitative approach analysts take to judging talent in the baseball market make my results informative to understanding the importance subjectivity takes in other important labor markets where such objective measures are absent (e.g. the hiring of corporate executives or high school teachers). Further study on subjectivity or on firm hiring preferences more generally can help labor economists understand a side of the hiring market which has long been ignored, and my paper provides the approach to do that.

Table 2: Explaining Salary

	(1)	(2)	(3)	(4)
WAR	17.286	5.072	4.896	5.029
	(0.450)	(0.771)	(.768)	(.766)
WAR Sq		3.156	3.146	3.128
		(0.170)	(.169)	(.168)
Age	-2.0048	-9.714	-10.2	-9.929
	(0.247)	(2.692)	(2.674)	(2.670)
Age Sq		0.122	0.130	0.126
		(0.0402)	(.0399)	(.0399)
Years	0.517	3.646	3.488	3.41
	(0.248)	(0.765)	(.761)	(.760)
Years Sq		-0.156	-0.151	-0.147
		(0.0362)	(.03603)	(.0361)
Position FE	X	X	X	X
Team Controls			X	X
Year FE				X
R-Squared	0.5425	0.6348	0.6426	0.6479
Adj. R-Squared	0.5403	0.6323	0.6385	0.6414
Residual Standard Error	19.66	17.58	17.43	17.36

Notes: This table presents coefficients from OLS regressions of player guaranteed salary on player and team characteristics, with guaranteed salary in 2006 dollars. Parantheses give robust standard errors. Data is on 1436 signed contracts between 2006 and 2016.

Table 3: MSE Estimates of Equation (3)

	(1)	(2)	(3)	(4)
WarAtt	0.4266	0.3627	0.0125	0.00258
TalOver	1.1116	0.0628	0.0106	0.0363
PosComp	-2.7402	-2.4573	-0.330	-0.0188
TalComp	-0.0774	-0.05702	-0.0121	0.0054
Re-Sign FE		X	X	X
Standalone			X	X
Square Terms				X
Score	0.299	0.334	0.425	0.468

Note: The coefficient units are $100k times variable unit.

3.0 The Role of Subjective Information in Hiring Decisions: An Experimental Study

3.1 Introduction

Labor markets are characterized by a number of subjective processes: situations in which different agents share the same information set, the same technology, and the same objectives, yet reach different conclusions. Examples include interviewing potential hires, writing worker performance reviews, determining performance pay, and delegating tasks amongst workers. In this paper, we use a simulated labor market to study the effect of subjective information in the hiring decision. We focus on the value of subjective information in hiring the right worker; the way in which subjective information interacts with objective information; and whether subjective information has heterogeneous effects across gender.

Understanding how subjective information is valued has the potential to explain several labor market puzzles. For example, subjectivity can explain how observationally equivalent workers receive such different labor market outcomes — a question which underlies the large literature on residual wage inequality (e.g. [36]). Another area of research is the growing importance of social and "soft" skills on labor outcomes (e.g. [18]); it's likely that traits such as work ethic, sociability, and leadership are judged more subjectively than attributes such as grade point average or ownership of an occupational license. Finally, subjectivity is important to understanding the role of "luck" in the labor market[1]. In many labor market instances, such as

[1]In 2018, *The Journal of Labor Economics* published an entire issue focused on the roles of productivity and luck in explaining the income distribution (see [42]).

choosing which candidate to hire, luck takes the form of a favorable (or unfavorable) subjective review.

We study the role of subjectivity in hiring decisions. The focus on hiring is due to its well-established importance in labor market outcomes. [64] and [20] both detail the importance of job switches in explaining wage development. The large literature using matched employee-employer data, beginning with [1], has consistently found that firm effects are important to wages. Finally, recent research has shown the long-run effects of early labor market outcomes, implying that hiring is especially important at the start of one's career[2]. That hiring matters so much to the outcomes of workers and firms is in sharp contrast to the lack of studies on the topic. In a recent *Handbook of Labor Economics* chapter, [55] highlight the overall lack of empirical studies on hiring, calling the process a "black box". How employers use subjective information to evaluate potential hires is a major part of that black box, as many of the signals employers rely on — interviews, references, resumes — are inherently subjective.

The major obstacle to understanding hiring — and especially subjectivity in hiring — is controlling for inter-firm differences in hiring and objectives. These differences make using observational data troublesome. For example, each firm employs unique screening and interview processes, leading to different subjective information collected across firms. Problems remain when looking within a firm, as different managers involved in a hiring decision might view a potential hire at different times, have different levels of input on the hiring decision, or have competing motivations. Any study using observational data would need to control for all of these firm differ-

[2]For example, [37] and [54] both show that graduating from college during a recession has negative and persistent effects on earnings, as graduates place in lower quality jobs initially; [3] find similar results, while also showing that such effects differ across college major.

ences, while also accounting for differences on the job-applicant side. For example, applicants self-select into which jobs to apply for, and even when applying for similar jobs, applicants may alter behavior based on experience or random factors (e.g. stress, poor sleep the night before an interview, or family issues).

We use an experimental labor market to avoid these challenges. To the best of our knowledge, we produce the first study which explicitly addresses subjective information in hiring. Our work has three main goals. First, we look at how hiring efficacy is affected by adding subjective information to a job applicant's profile. Second, we use treatments with and without subjective information to determine the trade-off between objective and subjective information in how managers value applicants. Third, we look at how the presence of subjective information affects gender discrimination in a setting where women are often discriminated against.

Our experiment consists of two types of sessions: worker sessions and manager sessions. In worker sessions, we collect worker profiles, which are akin to resumes. At the beginning of a worker session, subjects are asked to write a short statement about their ability to multiply numbers together, as if they were being interviewed for a job that requires this as a labor input. Next, subjects are asked to perform a multiplication task. After that, subjects engage in a related task, adding sets of numbers together. The short statement task provides us with a piece of subjective information for each worker profile, while the sum task provides us with objective information.

Manager sessions consist of two experiments. In the first one, we show subjects worker profiles we collected during the initial sessions, and we ask managers to value each worker. In the second experiment, two worker profiles are placed side-by-side, and managers are asked to select the one they believe scored higher on the multiplication task. In each manager experiment there are two treatments: one in which the worker's subjective information is included in the worker's profile, and one in which it is not. The use of a price list gives us a detailed measure of valuation, which we use to explore the objective-subjective trade-off; the use of side-by-side comparisons allows us to compare similar worker profiles, and see what effect subjective information has on anti-female bias.

We find that subjective information significantly increases the valuation managers assign to worker profiles and that it greatly reduces the weight placed on objective information, although it neither helps nor hurts the accuracy of valuation. These results are sensitive to manager gender, with female managers assigning significantly higher scores when subjective information is provided. We also find that subjective information strongly reduces the bias in favor of male workers in our setting, and that it causes experienced managers to make worse hiring decisions than they would without subjective information.

Firms — in practice — solicit and interpret a diverse set of subjective measures on job applicants. We show that subjective information has important implications for hiring in a very generic setting. Our results suggest that firms carefully consider the effectiveness and fairness of any subjective information they collect, and that academics be aware of subjectivity when studying hiring.

3.2 Related Literature

Our focus on information in hiring is most similar to [33], where they look at the introduction of analytic job testing of applicants across fifteen firms, in which job applicants are given a score by a machine based algorithm. A signal of the applicant's aptitude (green, yellow or red) is provided to a hiring manager by a third party firm, who then uses traditional signals of worker quality (e.g. interview and resume) to determine whether to hire a worker. The authors find that managers who hire against the test recommendation do worse on average.

While [33] obtain strong results regarding the efficacy of two different approaches to hiring, they do not open the black box of how managers made their hiring decisions. In particular, hiring managers don't know how the signal is constructed nor have any way to gauge how informative it is ex-ante, nor what kind of information is taken into account in the construction of the signal. Moreover, managers may have an incentive to selectively hire against the signal, in an attempt to prove to their employers that they are better than the third party's algorithm. By contrast, our experimental setting allows us to carefully control what information a manager sees. This allows us to test the hiring efficacy under two different information settings but also, importantly, allows us to say something about *how* managers reach their conclusions[3].

We include gender in our study, as there is good reason to believe subjective information interacts with demographics in a meaningful way. For example, [30] show that female candidates for symphony orchestras have their performance quality — which is subjective — judged more favorably when auditions are gender-blind. Another example comes from an ongoing lawsuit brought against Harvard University by a group representing Asian-American applicants[4]. Like many universities, Harvard asks applicants to submit personal essays and letters of recommendation, in addition to submitting objective measures such as grade point average and standardized test

[3]In addition, we also study a more common hiring setting than [33]. While it is true that more and more large employers have adopted analytic testing for job applicants, the majority of hiring is still done by small firms, employing less sophisticated approaches.

[4]See *Students for Fair Admissions, Inc. v. President and Fellows of Harvard College.*

scores. Amongst other complaints, the lawsuit argues that such subjective measures consistently harm Asian applicants[5]. By considering how gender discrimination is affected by subjective information, our work adds to a large literature using experimental labor markets to understand discrimination in hiring.

Several of these papers study discrimination against women in math related tasks and consider ways to reduce bias. [59] show that both male and female managers have a preference for men when hiring workers for a math task. They find that providing managers with a worker's previous score on the same exact task reduces, but does not eliminate, the bias. [10] show the same preference for male workers performing a math task, but show that it can be reduced when male and female workers are considered side-by-side. [17] study the preference for male workers using math and sports quizzes for performance tasks. Using a clever partition of workers by birth month, they decompose the preference for men, finding results that are consistent with statistical discrimination. Our work differs from these papers, in that it allows workers to submit a second piece of information to managers: their written statement. That we find evidence that this reduces discrimination against women creates the possibility for another way to limit preference for men.

While we believe our focus on subjectivity in hiring to be novel, subjectivity has been studied in other contexts. Studies of compensation show that considering subjective performance measures compresses compensation differences between workers[6]. Stock market trading has also been explained using subjectivity. While models of public information often lead to "no-trade" results (e.g. [51]), both [32] and

[5]The interaction between subjective information and an applicant's demographics relates to the legal literature on adverse impact. Under Title VII of the 1964 Civil Rights Act, it is illegal for a firm to request information which has a disparate impact on minority groups — regardless of intent — unless that firm can demonstrate that the information is "reasonably related" to the job. For a famous case regarding adverse impact, see *Griggs v. Duke Power Co.*

[6]For examples, see [52], [11], [12], and [7]

[38] allow traders to differ in how they interpret identical information, finding that such models help to explain trading behavior. Finally, the literature on "bounded rationality" has also allowed for agents to disagree over identical information. For example, [61] models the problem of a monopolist when consumers differ in their cognitive ability to process an identical signal[7].

The rest of the paper is structured as follows: Section 3 gives our experimental design; Section 4 presents our results; and Section 5 concludes.

3.3 Experimental Design

Our experiment consisted of two types of sessions: Worker Sessions and Manager Sessions. In Worker Sessions, subjects performed incentivized tasks. These tasks comprised a worker's "profile". In Manager Sessions, subjects were shown worker profiles and asked to forecast worker productivity. We now describe the Worker and Manager Sessions in detail. We refer to subjects in the Worker Sessions as "workers" and subjects in the Manager Sessions as "managers" throughout.

3.3.1 Worker Sessions

Worker Sessions were conducted in the Pittsburgh Experimental Economics Laboratory using zTree ([65]). Each session consisted of four tasks, one of which was

[7]While studies such as [61] and [38] allow agents to respond differently to information, their focus is on what effect these heterogeneous interpretations have on market equilibria. While our study looks at how subjectivity affects the hiring market, it also provides insight into how much disagreement there is amongst agents viewing the same information.

randomly chosen for payment. We conducted three such sessions, with thirteen subjects in each. We now discuss the tasks in detail, before giving an explanation for the setup.

The first task prompted workers to write a short statement about their ability to multiply numbers. Workers were given the following instructions for the task:

> You're trying to convince another person of your ability to multiply numbers. It may help to think of this as an interview question, where the job you're applying for requires you to multiply numbers. Therefore, feel free to mention anything from your academic background, work experience, personal interests etc. which you think will help you convey your ability to multiply numbers.

Workers were given ten minutes to complete task one. There were no restrictions regarding word count or content.

In the second task, workers had three minutes to solve as many multiplication problems as they could, with problems consisting of a random two digit number times a random one digit number.

Task three was used to evaluate each worker's written statement from task one. Each worker was presented with the written statement of the other 12 workers on their screen, one statement at a time. The worker was then asked to predict the other worker's performance on the multiplication task (task two) using **only** the statement from task one. After submitting a prediction, a worker was informed of the actual multiplication score associated with the statement they just evaluated. The order in which workers judge each other's subjective information was varied so as to avoid order effects.

In task four, workers had three minutes to solve as many sums of five, random two-digit numbers as they could ([53])[8].

[8] In tasks two and four (multiplication and addition, respectively) subjects were allowed to use scratch paper, but not allowed to use a calculator. No electronic devices were permitted at any

Before performing any of the tasks, workers answered a short demographic questionnaire which included age, gender, year in school, and major.[9]

For tasks two and four, workers were paid a piece rate of \$0.50 and \$1.00 per correct answer, respectively. In task 3, workers received \$3.00 per exact prediction and \$1 for each close prediction (within 3 points of the actual multiplication score). Finally, a worker's earnings for task one were given by \$0.50 times the average prediction other workers gave based on the worker's statement during task three. Because payment on tasks one and three were interdependent, workers were given a brief overview of the experiment prior to beginning task one.

Discussion

The first task provides us with a piece of subjective information for each worker. This will be part of the worker profile shown during Manager Sessions, with the idea being that different managers can read the same response to a prompt and interpret it differently. The prompt we use is similar to a worker being asked: "why do you think you'd be a good fit for this job?" — a frequent question in job applications and interviews.

Task three performs the important role of quantifying the quality of a worker's subjective information. We want to control for the information a worker's statement provides when presented to managers, in order to draw inferences about the importance of subjectivity in manager's choices. But how to compare the similarity between written responses is unclear. Our approach, of having other workers attempt to predict a worker's multiplication score using only her subjective information, is

point in the experiment.

[9] The questionnaire was conducted at the start of the experiment to prime truthfulness in Task 1 statements, particularly regarding a worker's major. Whether a result of this or not, no worker's statement in task one mentioned a major different from what was indicated in the questionnaire, nor were there any other inconsistencies between the two.

an attempt to reduce all of the information in that worker's written statement to a single value. We comment more on how task three is used when discussing results.

Having workers complete multiplication and addition tasks was done in order to tie our results to the aforementioned literature showing statistical discrimination against women in math-related tasks. The decision to use multiplication *and* addition — rather than just one task twice — comes from job applicants rarely having experience or training which perfectly aligns with a position's needs.

Finally, the answers to the questionnaire at the beginning of each session are akin to the type of information a manager would obtain from a worker's CV.

Out of the 39 subjects in the Worker Sessions, we discarded 3 for deviating from the statement's original intent; writing as to a fellow experimental participant instead of to a manager[10]. Out of the remaining 36 workers we removed anyone with a Sum Task score below 4 or above 9. These could be considered under-qualified and over-qualified workers, which would rarely apply to the job in the first place, or they would be immediately discarded or hired. This leaves us with a pool of 27 workers for the managers to evaluate.

3.3.2 Manager Sessions

After all Worker Sessions were run, we conducted Manager Sessions using Amazon Mechanical Turk. We implemented a 2x2 experimental design, in which the main treatment condition was whether subjective information was included in a worker's profile or not, while the other dimension varied with respect to evaluation method.

[10] For example, one subject ended their statement with the following sentence: 'I guess since its a theoretical job and maybe you wouldn't see me every day and I'm just ok at math you wouldn't give me a very high prediction but maybe there is some compassion in this cruel world so I don't really see a downside to you giving me a high prediction thank you <3.'

Managers were randomly assigned to either a Partial Information or Full Information treatment. In the **Partial Information treatment**, the worker profiles a manager was shown consisted of the worker's age, gender, year in school, and major, as well as the worker's score on the addition task[11]. In the **Full Information treatment**, the worker profiles a manager was shown consisted of all same information as in the Partial treatment, as well as the worker's written statement. Therefore, the two information treatments only differed with respect to whether a worker's subjective information was included or not.

We used an Individual Valuation and a Side by Side Comparison to have managers evaluate worker profiles. 159 managers participated in the **Individual Valuation treatment**, where they were shown 18 worker profiles each, one at a time. For each worker profile, a manager is required to submit a valuation for the worker using a modified price list mechanism[12]. Each round's payoff was therefore either a random offer, or — if the worker was hired — \$0.10 times the worker's performance in the multiplication task. Workers solved between 5 and 28 multiplication problems, leading to a price list ranging from \$0.50 to \$2.80, in ten-cent increments[13]. Once a valuation was submitted, the manager was informed of the worker's actual multiplication score, as well as what the manager's earnings for that round would be. The manager was then shown the next worker profile. Two randomly selected rounds, one from the first nine and one from the last nine, were chosen for payment.

In the **Side by Side Comparison**, each of the 496 managers saw five pairs

[11] Although the effects of age, year in school and major are not our main interests, we include them so as to make gender less salient.

[12] Our modified price list had the simplicity of a typical price list, but did not allow subjects to select multiple switch points: once a price was chosen, the rest of the list auto-completed according to said selection.

[13] Since real life hiring managers should have an idea of what range of worker performance they can expect, managers were provided with the overall max and min of worker performance on the multiplication and sum tasks.

of worker profiles, one at a time. For each pair, a manager had to try select the worker who performed better in the Multiplication Task. One of the five rounds was randomly selected for payment, and a correct selection in that round earned the manager \$2.00. Managers did not receive feedback after each round, and only learned whether they had made the correct selections after the experiment concluded. The pairs were selected such that the differential in the two workers' sum task scores was never more than 1. Furthermore, each manager saw: one pair consisting of two men; one pair consisting of two women; and three pairs consisting of a man and a woman. The three mixed gender pairs were constructed with possible score differentials on the sum task of +1 for the man, +1 for the woman, and even (zero differential in sum task score).

Discussion

In the Individual Valuation treatment we provide managers with feedback after each submitted valuation. This allowed us to check for learning effects between rounds. We did not provide feedback during the Side by Side Comparison; managers made just five binary selections in the Side by Side treatment, leaving us with insufficient power to test for learning.

As mentioned, the worker pairs in the Side by Side treatment were all within 1 point in their sum task score. The reason for this is that actual hiring decisions are often narrowed down to a choice between two competitive, similar applicants. If we allowed for pairs to consist of workers who scored very differently in the sum task, managers would overwhelmingly choose the candidate with the higher sum score; that choices are binary would make it impossible for us to see what effect subjective information had on preferences over workers.

3.4 Results

3.4.1 Workers

Table 4 shows summary statistics for the Worker Sessions. Subjects solved 14.49 multiplications and 5.82 sums on average. In Task 3 of our worker sessions, subjects were asked to use the statements of other workers to predict how many multiplications those workers solved. We use these predictions as our measure of the "quality" of a worker's statement. The average prediction the workers received from their peers based on the statement they provided in Task 1 was 14.96, with substantial heterogeneity in how statements were interpreted; for example the average predicted scores ranged from 10.08 to 21.92. This shows that worker statements were judged to be of varied quality.

Table 4: Worker Summary Statistics

Variable	**Mean**	**S.D.**	**Min**	**Max**
Multiplication Score	14.49	6.49	4	32
Sum Score	5.82	2.55	1	12
Words in Statement	139.00	53.71	16	229
Relevant Major	0.31	0.47		
Relevant Classes	0.49	0.51		
Avg. Predicted Score	14.96	2.72	10.08	21.92
N	39			

Notes: Relevant Major is a binary variable that indicates whether a worker mentions that their major is typically regarded as math intensive. Similarly, Relevant Classes indicates the mention of a worker having taken math intensive classes. Average Predicted Score is the average Multiplication prediction the workers received from their peers based on their Task 1 statement.

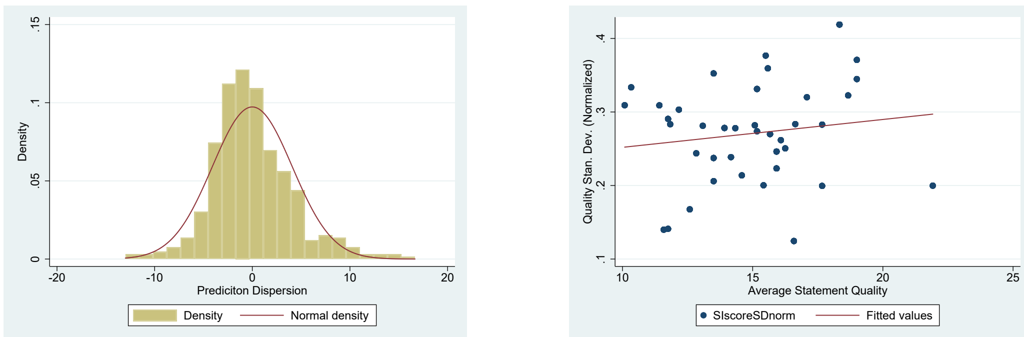
Moreover, there was substantial disagreement on the quality of each statement. Let i denote a worker, and $s_{i,j}$ denote the quality of worker i's statement, according to worker j. Therefore, if worker j reads worker i's statement, and predicts that i would correctly solve 10 multiplication problems, $s_{i,j} = 10$. We're interested in characterizing the distribution of $s_{i,j}$ in our worker sessions.

We begin by looking at the overall distribution of statement quality disagreement. For each statement, we calculate the mean quality score, and subtract it from the quality score given by a worker, i.e. $s_{i,j} - \bar{s}_i$ where $\bar{s}_i = \sum_{i \neq j} s_{i,j}$. We compute this differences for each of the 27 workers included in our manager sessions, giving us a measure of the disagreement over a worker's statement quality, net of the statement's mean quality. With each worker's statement being judged by 12 other workers, we compute a total of 324 such differences, and plot them in Figure 1 (left). The histogram shows that the distribution of subjectivity looks approximately normal[14], with mean 0 (by construction) and a standard deviation of 4.1. Superimposed on the histogram is the probability density function for a Normal distribution with the same mean and standard deviation. The Normal distribution fits our data well.

We now look at whether subjectivity differed across statements: were statements equally likely to generate disagreement, or were some more contentious than others? To do this, we calculate the standard deviation in quality for each statement, normalizing it by the statement's average quality. In Figure 1 (right) we plot these normalized standard deviations against average statement quality. The graph does not appear to show any relationship between the two (corr=0.148); there was as much disagreement for low quality statements as for high quality ones.

[14] We excluded one outlier from our data. One subject predicted that one of their peers solved 45 multiplication in 3 minutes, generating a disagreement of 26 points with respect to the mean prediction for that worker, and 13 points higher than the maximum score attained.

Figure 1: Statement Quality Disagreement



The histogram (left) shows the frequency distribution for our disagreement measure: $s_{i,i'} - \bar{s}_i$, across all worker statements. Superimposed on the histogram is the probability density function for a Normal distribution with the same mean and standard deviation as our disagreement measure ($\mu = 0$, $\sigma = 4.10$). In the scatter plot (right), we plot each statement's average quality against the standard deviation of its quality, where the standard deviation is normalized by the statement's average quality. The correlation between the two is 0.148

Next, we look at whether written statements are informative of worker multiplication scores. We do this in Table 5. Subjects in the worker sessions seem to be able to correctly use the subjective information to make better than average predictions of others' performance. In an attempt to understand this, we look at the effect of mentioning a major stereotypically considered as math intensive (STEM, Economics, or Business), mentioning having taken math intensive courses, and the length of the statement. This can be considered as objective information included in the statement. Once we control for these variables, we find that the statement is no longer informative of worker multiplication score. In particular, the mention of a math intensive major is the only variable that is significantly predictive of a higher score in the multiplication task.

In the manager sessions, all worker profiles include their major. Therefore we can safely use our 'statement quality' measure in the following section since any mention of a relevant major in the statement should be irrelevant for the managers.

3.4.2 Managers

We first look at the Individual Valuation treatment. In this setup, we obtain a precise valuation for worker profiles. This allows us to study the trade-off between objective and subjective information, as well as any differential valuation due to a worker's gender. We begin by looking at manager valuation errors, and whether subjective information helps reduce valuation errors.

We find that manager predictive ability is indistinguishable across information treatments. We define a manager's error to be the absolute difference between a worker's actual multiplication score and the manager's predicted score for that worker, where the manager's predicted score is given from the price list valuation. In the partial information treatment, manager's make an average error of 5.36, while average error in the full information treatment is 5.40, with the difference being statistically insignificant (p-value 0.79). While average errors under either information treatment are indistinguishable, managers in both treatments would have done significantly better by choosing the midpoint of the price list for every worker (average error 4.86).

Table 5: Predictiveness of Subjective information Score

Variable	**(1)**	**(2)**
Predicted Score	0.735*	0.291
	(0.374)	(0.398)
Relevant Major		4.680*
		(2.475)
Relevant Classes		-0.685
		(2.041)
Words in Statement		-0.0196
		(0.0207)
Constant	3.493	11.74*
	(5.686)	(6.678)
Observations	39	39

Standard errors in parentheses [* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$]

Note: Both specifications are ordinary least squares regressions. The dependent variable is the score in the Multiplication task.

Table 6: Information Effects in Worker Valuation

Variable	**(1)**	**(2)**	**(3)**
Sum Task Score	1.287***	1.226***	1.060***
	(0.0590)	(0.0610)	(0.0464)
Full Information	3.763***	3.735***	1.615***
	(0.712)	(0.708)	(0.464)
Sum Score * Full Info	-0.371***	-0.367***	
	(0.0882)	(0.0880)	
Worker is Male		0.601***	0.860***
		(0.154)	(0.202)
Male * Full Info			-0.561*
			(0.294)
Constant	5.178***	5.400***	6.361***
	(0.477)	(0.478)	(0.414)
Observations	2862	2862	2862

Robust Standard Errors in parentheses [$^{*}\, p < 0.1$, $^{**}\, p < 0.05$, $^{***}\, p < 0.01$]

Note: The results are from a Generalized Least Squares regression where the dependent variable is the worker valuation given by the managers. Full Information is a binary variable indicating if a manager saw the subjective information as part of the worker profiles, and this is interacted with the other two variables of interest in the regressions.

That both information treatments produced similarly large errors is not due to disregard for the information in the worker profiles. Table 6 regresses worker profile valuation on profile components. Columns 1 and 2 show that an extra point in the Sum Task score lead managers to increase their prediction of worker performance by 1.2 points, which is significant. The table also shows that managers valued workers more highly in the full information treatment, and that the presence of subjective information lead to managers putting less weight on objective information (Sum Task score). That workers were given higher valuations under full information is robust even at the individual worker level: 25 out of the 27 workers received higher average predictions when subjective information was provided[15]. Figure 2 plots the distribution of values across information treatments. It shows that the presence of subjective information pushes the entire distribution to the right.

We also find a small but significant gender effect in favor of male workers (Table 6 Column 2). After controlling for their Sum Task score, men are predicted to solve 0.60 more multiplications than women (4.3% of the mean valuation). This discrimination is severely mitigated when providing subjective information (Worker is Male - Male*Full Info): the valuation-gap (i.e. wage-gap) decreases by 65%.

This mitigating effect of subjective information on gender discrimination could be caused by women writing better statements than men. We check for this by comparing the results from Task 3 of our worker sessions. In Task 3, workers were asked to predict other workers' multiplication scores using only their written statements. We find that the predictions male and female workers receive are indistinguishable. Male workers are predicted to score 14.38 in the Multiplication Task, while female workers are predicted to score 14.65 (p=0.94)[16]. Given that workers didn't see the

[15]These differences were significant at the 10% level in 15 out of the 25 cases and to the 5% level for 11 of those.

[16]These values correspond to the 27 worker statements that were shown to the managers. We

gender of who they were evaluating, this shows that the shrinking of the gender gap cannot be attributed to women writing better statements.

To further understand these results, we look at manager gender. We find that female managers, on average, increase their valuation by 2.49 points (p-value 0.001) when the subjective statement is included in a worker's profile, while men increased their valuation by just 0.58 (p-value 0.363).

obtain similar results on the entire worker sample.

Figure 2: Valuation by Information Treatment

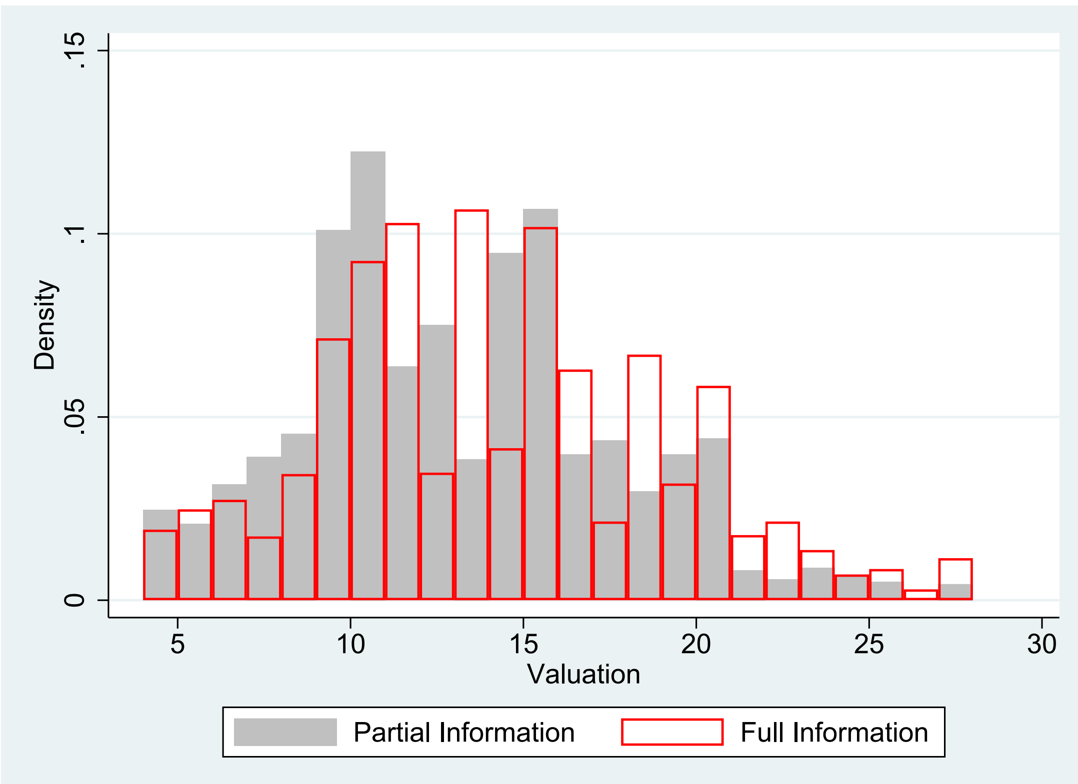


Table 7: Information use by Manager Gender

Variable	**Female Man. Partial Info**	**Male Man. Partial Info**	**Female Man. Full Info**	**Male Man. Full Info**
Sum Task Score	0.972***	1.439***	0.590***	0.989***
	(0.0866)	(0.0785)	(0.132)	(0.0864)
Worker is Male	0.477*	0.881***	0.571	0.632**
	(0.286)	(0.260)	(0.435)	(0.290)
Predicted			0.0600	0.192***
			(0.0754)	(0.0516)
Constant	6.518***	4.448***	10.47***	5.149***
	(0.654)	(0.626)	(1.414)	(0.953)
Observations	756	828	522	756

Robust Standard Errors in parentheses [* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$]

Note: These results are from Generalized Least Squares regressions where the dependent variable is the worker valuation given by the managers. The first two columns include only the managers that did not see the subjective statement as part of the worker profiles, while the last two correspond to the Full Information treatment.

Table 7 refines the analysis of Table 6 to account for manager gender. For female managers, the increase in valuation under full information is irrespective of statement quality; the subjective information score we constructed from the worker sessions doesn't significantly explain how female managers value workers (row three in Table 7). We do find that male managers positively react to a statement's quality. Table 4 also shows that female managers put less weight on the sum task score than their male counterparts, while both genders decrease this weight when subjective information is provided. The results also show that gender-wage gap from Table 6 is primarily the result of male managers.

We now turn to results from the Side by Side comparison (SbS). We begin by looking at manager predictive ability and how it's affected by information treatment. Overall, we find that managers don't do better than chance, picking the best worker just over 50% of the time: 50.7% with subjective information and 50.1% without it (p-value 0.746). However, as in the individual valuation treatment, this isn't due manager inattention. We find that managers choose the worker profile with the higher sum task score 69.7% of the time, and that this is sensitive to the information treatment. As shown in Figure 3, when subjective information is provided, the profile with the higher sum score is chosen 65.7% of the time, compared with 73.6% of the time when no subjective information is provided (p-value 0.001).

The small gender discrimination effect we found in the Individual Valuation treatment translates to more drastic results when workers are compared side by side.

Overall, male workers were chosen 55.4% of the time in male-female pairings. This bias is especially strong when men have the higher sum task score: male workers are chosen 48% of the time when both candidates have the same sum task score, but 81% of the time when they have the higher sum score. By comparison, women are only chosen 62% of the time when they have the better objective credentials.

This preference for male workers is completely removed when subjective information is provided. Under full information, male workers are chosen just 51.5% of the time, which is not statistically different from 50% (p-value 0.418). By comparison, male workers are chosen 59.3% of the time when subjective information is not included. The pairings in which male workers have a higher sum score is responsible for removing all of preference for male profiles when subjective information is added; males go from being chosen 91.6% of the time under partial information to being chosen just 69.5% of the time under full information. As stated before, this cannot be attributed to women writing better statements than men.

Our experiment is also idoneous to test the hypothesis proposed in [10]. In that paper, they find evidence that discrimination against a group perceived to have a disadvantage in a certain task is mitigated when workers are compared side by side. They use a mathematical task as the stereotypically male task, and find that a side by side comparison of worker profiles mitigates the preference for male workers they find when workers were valued individually. Considering only the 30 pairs of workers used in our side by side treatment, we constructed a counterfactual side by side choice using the managers' predictions from our individual valuation treatment; assuming that if presented with a pair, each manager would have selected the worker to which they assigned the higher valuation. In contrast to Bohnet et al. (2016). Overall, we find no significant difference in the proportions of male workers chosen from male-female pairs between the actual and the counterfactual (55.4% vs 57.2%, p=0.40). Nevertheless, we do find some evidence for said effect in our Full Information Treatment; while male workers are selected 51.5% of the time when compared Side by Side, they would have been chosen 56.3% of the time according to their valuations in the IV treatment (p-value 0.107).

A curious result arises from one question we asked our managers: 'Do you have any work experience in Human Resources?'. Though only a small number of managers answered this question in the affirmative[17], we find significant differences between 'experienced' and 'inexperienced' managers in their predictive accuracy and in how they judge subjective information.

Managers with HR experience do significantly better overall at picking the right candidate in the Side by Side treatment, but they are only more successful than their inexperienced counterparts when judging profiles without subjective information. When evaluating workers individually, HR and non-HR managers are equally accurate overall, but subjective information makes experienced managers' predictions significantly worse as shown in Figure 4.

3.5 Conclusion

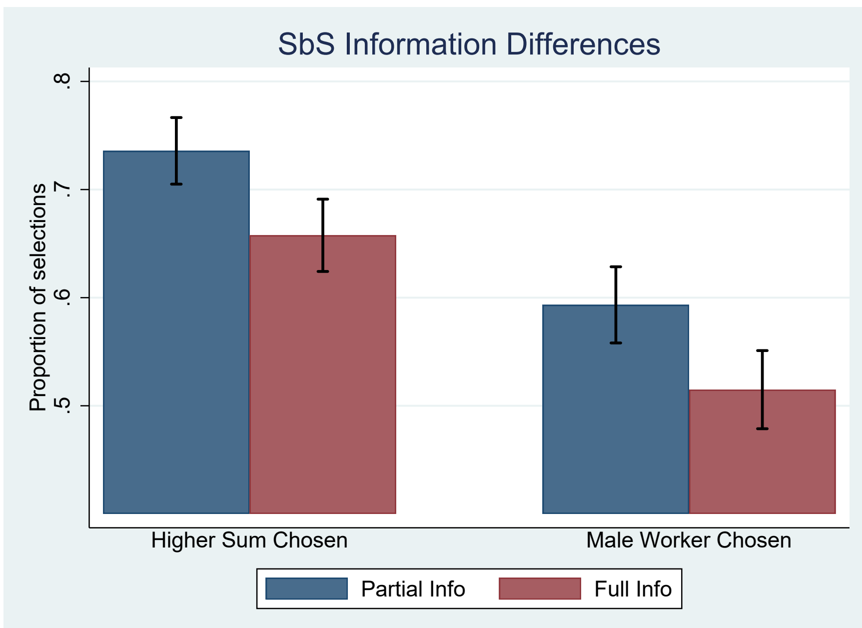
In this paper we consider the role subjective information can play in a hiring setting. Using an experimental labor market, we're able to isolate the information channels that make studying subjectivity so difficult in field data. Our work is motivated by an interest in the efficacy of subjective information, as well as how it affects the weight managers give to objective measures of ability and to gender.

What we find is that subjective information changes the overall valuation managers assign to worker profiles, although it doesn't lead to better or worse hiring results on average. Furthermore, these results are sensitive to manager gender. Our strongest findings come from the effect of subjective information on gender discrimination. There we find that subjective information strongly reduces the well-known bias in favor of male workers when it comes to hiring for math-related tasks and

[17] 20 of 159 in IV and 72 of 496 in SbS

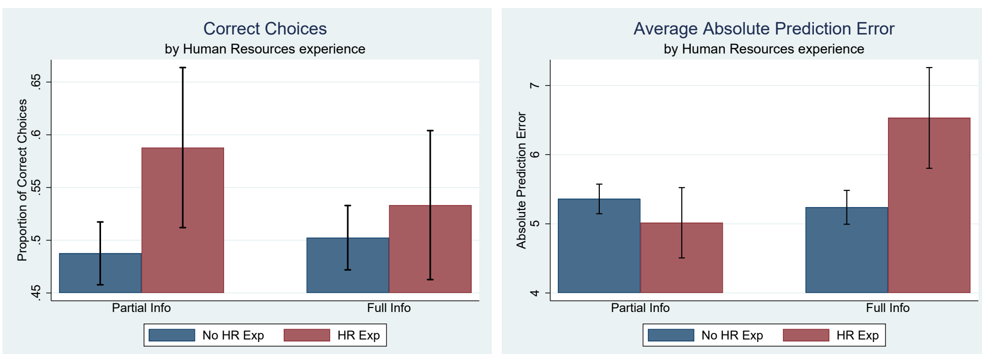
it's unrelated to any potential gender differences in verbal ability. This implies that subjective information — and the type of subjective information a hiring manager collects — are potentially useful tools in combating discrimination. It also implies that hiring tasks are sensitive to differences in subjective information provided — something that researchers should keep in mind in future work. Finally, we find that experienced workers do much worse when the worker profiles they are evaluating include subjective information, suggesting that collecting and evaluating this type of signals may be most prejudicial where it matters most.

Figure 3: Information Effects in Side by Side Comparison



The figure shows the proportion of selections of the worker with the higher Sum Task Score (left) and the male worker (right). Vertical segments represent 95% confidence intervals.

Figure 4: Valuation Error by Experience



The left graph shows the proportion of correct choices by manager experience and Information in the Side by Side treatment. On the right, we see the average prediction error in the Valuation Treatment for the same populations. Vertical segments represent 95% confidence intervals.

4.0 How Subjective are Statements in a Hiring Experiment?

4.1 Introduction

Labor economics is filled with studies concerning worker-firm matching, the results of which shape economists' understanding of important issues, such as unemployment and inequality.

Paramount in studying worker-firm sorting is attempting to control for unobserved heterogeneity on the two sides of the market. Traditionally, such models have attributed much of the sorting between workers and firms that is not explained by observable covariates to unobserved ability on the part of the worker and unobserved productivity on the part of the firm. These models generally assume that homogeneous firms (in terms of technology, productivity, size, etc) agree about a particular job applicant's value to their respective firms.[1]

Forcing homogeneous firms to agree about an applicant's quality is at odds with the prevalence that subjective information plays in the hiring market. When I say subjective information, I am referring to any piece of information for which two identical agents (in terms of technologies, preferences, characteristics, etc) can both view the information, and both come away with different opinions as to its quality.

For example, consider a situation in which two consumers view the same television commercial for a new product. Even after watching the same advertisement, it is unlikely that the two consumers completely agree with one another as to the new product's quality. In this way, when making their consumption choices, the new product's quality enters the decision processes of the two consumers at different

[1] For example, see [58].

levels; i.e. the quality-dosage that the consumers have been treated with, is idiosyncratic, and unobservable to the researcher.

In the labor market, subjective information is collected throughout the hiring process. Job applicants submit reference letters, which can give readers differing impressions; they answer interview questions which have no objective answer; they eat lunch with potential co-workers who must judge what the candidate would be like to work with; etc.

Presumably firms use these subjective pieces of information when making hiring decisions, as their collection is costly. This then begs the question, how disagreeable are job candidates along subjective dimensions? If the answer is very disagreeable, then this has important implications for how much of worker-firm sorting is based on unobserved ability and productivity – which are concrete – and how much is based on unobserved differences in subjective impressions.

This paper attempts to answer that question by looking at a well-defined hiring market, with one well-defined piece of subjective information. The hiring market is an experimental one, in which job applicants and hiring managers complete tasks and make decisions in a lab setting. Both the setting and the economic data for this project comes from [6]. In that paper, hiring managers evaluate worker profiles, which can be thought of as akin to resumes. Each worker profile contains a set of objective covariates (i.e. covariates for which all managers are in agreement over valuation) which managers use to project a worker's productivity.

Half of the managers are assigned to a treatment in which the worker profiles they view include an additional piece of information: a written essay from the worker. The written essay is similar to the answer to a common interview question: "tell me why you think you would be good at this job?" Because the essay's quality is impossible to quantify, it serves as a piece of subjective information that managers can use in their valuation decisions.

This paper goes beyond [6] by attempting to quantify the disagreement amongst managers as to the quality of each worker's statement. This is challenging, as a manager's valuation of a worker's statement is a latent variable.

To do this, I make use of two unique features present in [6]: a proxy variable for essay quality, and the treatment/control setting, in which half of all managers did not view the essay. The proxy variable allows me to represent the regressions of [6] in terms of a Berkson error model. This leads to consistent and unbiased estimation results for the regression model. These consistent estimates are important in the second stage of my estimation, as I rely on a deconvolution approach to recovering the distribution for subjectivity (i.e. the distribution of disagreement over essay quality).

The deconvolution approach requires me to make an assumption about the error distribution in the manager regression model. I assume that managers who see a worker's essay and those who do not see an essay, have regression models with the same error term distribution. In this way, I use residuals from those managers who did not view essays, and separate them out from the residuals of managers who viewed essays. This leaves the distribution for subjectivity, subject to a scaling parameter.

I find is that the distribution of subjectivity is bell-shaped and centered around zero, but that it can have considerable impact in how a written statement's quality is judged. For example, for a worker statement which is judged to be of median quality, the difference in receiving a subjectivity shock at the 25th or 75th percentile of the subjectivity distribution amounts to a roughly 30% difference in how managers interpret the quality of the statement. While the real-world importance of subjectivity is likely to depend on the setting and the sort of subjective information collected, the fact that answers to a simple and common question, "why do you think you would be good for this job?" can elicit such sharp disagreements implies that subjectivity is potentially important to how workers and firms sort.

In addition to showing the potential importance and level of disagreement that subjective information induces, this paper – so far as I know – is the first in the economics literature which combines a lab setting with deconvolution to recover the distribution of a latent variable. The deconvolution problem that I solve is well-known in the statistics literature on measurement error, and I show how it can easily be extended to experimental economic settings. As economists become more and more interested in unobserved heterogeneity and the effects of latent variables, the experimental setting, with its ability to dictate treatments, is conducive for estimating other latent densities of interest.

Related Work

[5] and [6] are two papers which look at how subjective information impacts hiring decisions. [5] uses observational data from professional baseball contracting to model worker-firm matching and salary negotiations. He finds that subjective information is pivotal for five percent of the labor market, in that their employment status would change from employed to unemployed, or vice-versa, depending on how their subjective information is interpreted. [6] simulate worker hiring in a lab experiment. They study how the provision of subjective information affects: worker valuations; the weight given to objective measures of worker performance; and gender-biases.

While the two aforementioned papers study the effect that subjective information provision has, neither is able to define one piece of subjective information and to characterize how much disagreement managers have over it.

Outline

The rest of this paper is structured as such. Section 2 describes the setting and data collected in Azic and Lamé. Section 3 introduces the economic model from which subjectivity will be recovered. Section 4 provides the deconvolution and estimation results for subjectivity. Section 5 concludes.

4.2 Background on Azic and Lamé (2018)

The model and application in this paper are based on the economic setting and data collected in [6]. Azic and Lamé simulate a hiring market in a lab experiment, where participants are randomly assigned the role of workers or managers. Sessions involving workers and managers are conducted separately. The goal of worker sessions is to create worker-profiles, which are akin to resumes. These worker-profiles are then shown to managers during a manager session. The job of the manager is to best guess what a worker's productivity is, given his or her worker-profile.

Worker sessions occur in an experimental lab. At the start of the lab session, participants (i.e. "workers") are asked to compose a short written essay about their ability to multiply numbers correctly. Workers are told to think of the essay as if they were applying for a job which requires them to multiply numbers correctly, and that they should feel free to mention anything from their background or interests which would convey their proficiency in multiplication (e.g. education, previous work experience, etc.).

After composing their essays, workers perform a multiplication task, in which they are given a set amount of time to solve as many multiplication problems as possible. This is then followed with a sum task, similar to the multiplication task, but testing addition rather than multiplication.

Next is a task which judges the quality of the statements written in the first task.

In this task, each worker in the room sees the statements of each other worker in the room, one at a time. Using only the statement of a fellow worker – and knowing nothing else about that worker – the worker is asked to guess how many multiplication problems the worker who wrote the statement solved. This provides a measure of a written statement's "quality". These measures will be important later, as they will serve as a proxy for a manager's subjective evaluation of a worker's statement. After each worker has evaluated the scores of his or her fellow workers, the session ends with a demographic survey, asking questions about gender, race, educational background, age, etc.

At the end of the worker sessions, a worker-profile has been created. The short statement that workers write serves as a piece of subjective information in the worker-profile, as its quality is idiosyncratic from the perspective of each manager who views the statement. By contrast, demographics and sum task scores serve as objective measures, as every manager will agree upon their values.[2]

During manager sessions, participants playing the role of managers are assigned to one of two treatments: Information or No Information. In both treatments, the manager sees a worker's profile, the only difference being that a worker's profile contains the subjective statement in the Information setting only. In both treatments, a manager sees a worker profile, one at a time, and is prompted to submit his valuation on that worker, where the valuation is solicited via a price list mechanism. The valuation solicited can be thought of as analogous to the expected number of

[2]This is not to say that every manager will value, or weight, objective measures of sum score and demographics in the same exact way, only that they will agree upon the information regarding sum score and demographics, i.e. in the parlance of linear regression, managers agree about the "x values", but not necessarily the "β values." In section 5 I consider the case where the βs may differ.

multiplication problems a worker solved, conditional on that worker's profile.[3]

In the remaining sections of the paper, the goal is to use the setting in [6] – which is typical of most experimental hiring markets – to try to uncover how subjective the worker statements were. I ask the question, is there a way to use manager valuations in order to determine how much disagreement over the quality of worker statements there was?

To do this, I make use of the proxy variable created during worker sessions, in which each worker was asked to evaluate the statement quality of his or her fellow worker by – using only that statement – trying to guess the other worker's multiplication score. Each worker session consisted of 13 workers, which means that each worker's statement was rated by 12 other workers. Azic and Lamé take those 12 scores and average them to provide their proxy variable for a worker's statement quality.

Azic and Lamé also analyze the worker statements and look into how each statement's quality was disagreed upon amongst the 12 workers who reviewed it. They establish several descriptive facts about how worker statements were reviewed. One fact, which will prove important in the subsequent section, is that the average quality score a statement was given was *not* correlated with how disagreeable workers found the statement to be. In the next section, I will make the same assumption from the perspective of managers: the level of disagreement about a statement amongst managers is not correlated with the statement's underlying quality proxy.

As mentioned, [6] look at the level of disagreement that workers had over a statement. They find that it is approximately Normal-shaped. However, this is an incomplete measure for subjectivity. Namely, the setting in which workers review

[3]The details of the price list mechanism are given in [6] and follow well-established practices in experimental economics that have been shown to approximate this definition.

the statements of other workers is fundamentally different than the setting in which managers review worker-profiles. Therefore, while the worker sessions provides a necessary proxy for statement quality and some insights into the relationship between statement quality and disagreeableness, it is necessary to measure subjectivity from the manager valuations.

4.2.1 Insights About Subjectivity

In the next section, I show how the proxy makes the results in [6] consistent, and how it can be used to recover the distribution of statement subjectivity.

4.3 Model

In this section I lay out a manager valuation model corresponding to the economic setting described in section 2. I then provide a set of assumptions under which the model can be used to determine the degree of subjectivity in written statements, and sketch the deconvolution argument that will be used in Section 4.

Let $i = 1, ..., I$ denote workers, and $j = 1, ..., J$ denote managers. Each manager sees a subset of the I worker profiles, and is tasked with assigning a valuation to each worker-profile shown. Managers are randomly assigned to one of two treatments: *Information* or *No Information.*

In both information settings, the manager sees a worker profile that consists of the worker's objective characteristics – including sum task score and demographic information. Let X_i denote the objective variables in the profile of worker i. In the Information setting, the manager sees an additional piece of information: the worker's written statement. Denote the quality of worker i's written statement as $Q_{i,j}$; that the subscript includes j is meant to capture the subjectivity involved with evaluating a written statement's quality. Managers in the No Information treatment do not see a worker's statement.

Given a worker's profile, it is assumed that managers use a simple linear model to assign values to a worker. For the full information treatment, let $V_{i,j}$ denote the valuation that manager j assigns after viewing worker i's profile $(X_i, Q_{i,j})$. Then the full information model is:

$$V_{i,j} = \beta_0 + X_i\beta_1 + Q_{i,j}\beta_2 + \mu_{i,j}, \tag{4.1}$$

where $(\beta_0, \beta_1, \beta_2)$ represent coefficients managers use when judging a worker profile, and $\mu_{i,j}$ represents an uncorrelated error term with mean zero and variance σ^2_μ. For the partial information treatment, let $\widetilde{V}_{i,j}$ represent the valuation that manager j assigns to worker i, and the partial information model to be given by:

$$\widetilde{V}_{i,j} = \alpha_0 + X_i\alpha_1 + \epsilon_{i,j}, \tag{4.2}$$

where (α_0, α_1) represent coefficients managers use when judging a worker profile, and $\epsilon_{i,j}$ represents an uncorrelated error term with mean zero and variance σ^2_ϵ.

In equation (1), $Q_{i,j}$ is not visible to the researcher, i.e. the researcher doesn't know how highly manager j thought of worker i's statement. Omitting $Q_{i,j}$ would lead to omitted variable bias (due to the correlation between $Q_{i,j}$ and X_i). To avoid this, I use the proxy variable $\bar{P}_i$, which Section 2.1 described and which comes from the worker's sessions, in place of $Q_{i,j}$.

Assumption 1: *The relationship between the proxy variable,* $\bar{P}_i$, and manager j's evaluation of worker i's statement, $Q_{i,j}$, is given by the relationship $Q_{i,j} = \bar{P}_i + S_{i,j}$, with $E[S_{i,j}|X_i, \bar{P}_i] = 0$.

The assumed relationship between the proxy and the actual subjective value takes the form of *Berkson measurement error* [9].[4] Note that Berkson errors are distinct from the more common, *classical* measurement error, in which the researcher observes the convolution of the true covariate and a measurement error. Whereas classical measurement error implies that the researcher observes more variation in a covariate than there really is, Berkson measurement error implies the opposite: the actual covariate ($Q_{i,j}$) has more variation than the proxy being used ($\bar{P}_i$).

What's important for my study – and for the validity of the results in [6] – is that under Berkson measurement error, consistent estimation results can still be had. This is explained in Result 1.

Result 1: *In equation (1), replacing $Q_{i,j}$ with its proxy, $\bar{P}_i$, gives unbiased and consistent least squares estimates.*

To see Result 1, substitute the equation for proxy into model given in equation (1):

$$V_{i,j} = \beta_0 + X_i\beta_1 + (\bar{P}_i + S_{i,j})\beta_2 + \mu_{i,j}$$

$$= \beta_0 + X_i\beta_1 + \bar{P}_i\beta_2 + (S_{i,j}\beta_2 + \mu_{i,j}). \quad (4.3)$$

[4]Berkson errors are common in the epidemiological studies, in which a subject's exposure to a harmful environmental agent cannot be directly measured, but for which the researcher may have knowledge of the concentration of over a large area.

By assumption, $E[S_{i,j}|X_i, \bar{P}_i] = 0$, and the results based on (3) are unbiased estimators for the beta coefficients.

The goal of this paper is to recover the distribution of the random variable $S_{i,j}$, as this will tell us what disagreement over subjective statement quality looks like. To do this, I make the assumption that the error terms in the Information and No Information treatments have the same distribution.

Assumption 2: *For the models in equations (1) and (2), $\mu \overset{d}{\sim} \epsilon$.*

Assumption 2 appears strong, however most all solutions to measurement error issues require some auxillary information (see [14], Chapter 2). In my case, what is important is that I impose as few assumptions as possible on the distribution S itself. Furthermore, the assumption that $\mu \overset{d}{\sim} \epsilon$ is implicit in most experimental designs, an exception being designs where error are clustered by treatment group.[5] Assumption 2 makes it possible to recover the distribution of S via deconvolution.

Result 2: *Under Assumptions 1 and 2, the distribution of subjectivity can be recovered from a deconvolution argument.*

To see this, let hats denote in sample estimates from a least squares regression. By Result 1, a regression of equation (1) gives consistent estimators $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$. Using these estimates, the residuals from the regression provide a consistent estimate for $(S\beta_2 + \epsilon)$ (see equation (3)). Under the assumption that $\mu \overset{d}{\sim} \epsilon$, deconvoluting the residuals from the No Information regression model from those of the Informa-

[5]Note that [6] randomly assign managers to treatments. Therefore it is unclear why errors should be clustered at the treatment level.

tion regression model, leaves a consistent estimate of the distribution $S\beta_2$. Having a consistent estimate of β_2. this can be scaled by $1/\hat{\beta}_2$, leaving a consistent estimate for the distribution of S.

I now detail how the deconvolution is accomplished, and provide estimates for the distribution of S.

4.4 Estimation

Throughout this section, I use the following notation. Define $\widetilde{S} \equiv S\beta_2$ and $W \equiv \widetilde{S} + \epsilon$. Let $f_{\widetilde{S}}$ represent the density function of random variable $\widetilde{S}$; f_W represent the density of W; and f_ϵ represent the density of ϵ.

The random variables $\widetilde{S}$ and ϵ are independent of one another. Therefore, their sum can be written as the convolution of their densities:

$$f_W(w) = \int f_{\widetilde{S}}(\widetilde{s}) f_\epsilon(w - \widetilde{s}) d\widetilde{s}. \tag{4.4}$$

For now, assume that f_W and f_ϵ are both perfectly known, and do not require estimation. Let ϕ_W denote the characteristic function of W and ϕ_ϵ denote the characteristic function of ϵ, where $\phi_W(t) = \int e^{itw} f_W(w) dw$ and $\phi_\epsilon(t) = \int e^{itw} f_\epsilon(\epsilon) d\epsilon$. Then, $f_{\widetilde{s}}$ can be recovered by a Fourier transformation:

$$f_{\widetilde{S}}(\widetilde{s}) = \frac{1}{2\pi} \int e^{-it\widetilde{s}} \frac{\phi_W(t)}{\phi_\epsilon(t)} dt. \tag{4.5}$$

While equation (5) works in principle, the difficulty comes from the fact that f_W and f_ϵ are unknown, and therefore must be estimated. Kernel density estimation of f_W will produce an estimated characteristic function, $\hat{\phi}_W$, which can be plugged into equation (5). However, for most commonly used kernel functions – including Gaussian and Epanechnikov – the Fourier transform is not well-defined, leading to integrals that cannot be computed in equation (5). [63] detailed this for the problem where f_ϵ was known and only f_W had to be estimated. They gave examples of kernels for which the Fourier inversion is well-defined, one of which is the sinc kernel, which I use in estimation. [19] brought up an additional challenge relating to equation (5) for the setting where both f_W and f_ϵ had to be estimated; in low density regions, the ratio of ϕ_W/ϕ_ϵ will produce highly unlikely estimates for $f_{\widetilde{S}}$, which are extremely sensitive to outliers.

In this paper, I follow the approach of [63]. [63] develop a kernel density deconvolution estimator which convolutes the characteristic function of the kernel with the empirical characteristic function based on the sample of $\{W_i\}_{i=1}^{I}$, and replaces ϕ_W in equation (5) with this estimate. I do this while using the sinc kernel, $\frac{sin(x)}{\pi x}$, which has a well-defined Fourier transform that makes integration possible. To compute density estimates, I use the R package *kerdec* ([8]), which makes use of fast-Fourier transform software to solve integrals in a matter of minutes.

In addition to the issues posed by integration and by estimation at the tails, there are two other estimation challenges – one fundamental and the other specific to my application.

The first challenge is that the kernel deconvolution estimator has an extremely slow rate of convergence (see [15]; [63]; [19]; and [34]. While the estimator's rate of convergence depends on the curvature properties of f_ϵ and $f_\S$, convergence happens at a logarithmically slow rate, with many common settings haveing rates of convergence bounded by $(log\, n)^{-1}$ or $(log\, n)^{-2}$, where n is the sample size.[6]

[6]For example, if f_ϵ is Normally distributed and f_S has k bounded derivatives, [15] show that

Another issue with the estimator, which is germane in my setting, is its ability to separate a signal when its noise is large. Figure 1 plots densities for the residuals from the Information and No Information settings in Azic and Lamé. In agreement with Assumption 2, the residuals from the Info setting are more dispersed than those in the NoInfo setting. However, the overall difference between the two densities appears to be small. In this way, much of the variability in the Info residuals is accounted for by the NoInfo residuals, i.e. the remaining "signal" – representing subjectivity – is weak relative to the noise of the error distribution.

While the residuals from the Info and NoInfo settings look similar in Figure 1, the distributions are, in fact, statistically distinct. I check for this using two tests, based on [44] and [47]. [44] uses a test statistic based on an integrated square difference of the kernel densities, while [47] is based on a normalized Hellinger-Bhattacharya-Matusita entropy metric.[7] In both cases, the null hypothesis that the two residual distributions are the same is rejected below the 0.1% level. Therefore, while my deconvolution estimation may perform better in a setting where the difference between f_W and f_ϵ is more stark, the statistically significant difference in the Figure 1 distributions allays concerns about the estimator.

Figure 1 also shows that the residuals from the NoInfo setting, which correspond to the error distribution f_ϵ, look approximately Normal. Deconvoluting with a Normal error distribution is undesirable, as its smoothness retards the estimator's rate of convergence. The intuition for this is that it is easier to remove the noise from an error if it has a strong peak in its distribution than if the error's distribution is multimodal. For example, although a LaPlace distribution has the same general

---

the fastest rate of convergence any estimator can have is $(log\ n)^{-k/2}$.

[7]Both papers derive test statistics based on kernel density distributions. Software in the R package "np" was used to conduct the hypotheses tests.

shape as a Normal distribution, the kernel deconvolution estimator has a better rate of convergence when the error's distribution is the former ([63]).

Because of these challenges, I estimate the density of $f_{\widetilde{S}}$ under three different approaches, and look for whether they agree. The first two estimate f_ϵ parametrically, using the residual sample from the NoInfo setting. One estimate assumes that f_ϵ is distributed Normally, while the other assumes f_ϵ comes from a LaPlace distribution. The third approach estimates f_ϵ nonparametrically. The resulting densities are plotted in Figure 2.

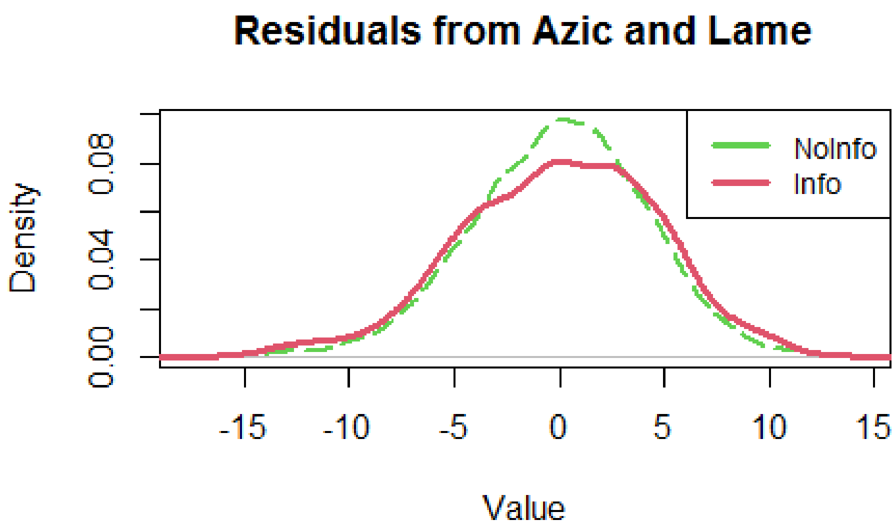
The densities in Figure 2 appear to agree on a general shape for f_S, with a single peak, and shape somewhat near Normal.[8] All three densities appear to be centered slightly to the right of zero, and have small disagreement as to where their peaks are. Although all three produce some minor oscillations, the LaPlace appears to be the most stable. Note that all three produce small regions in which the density is negative, as mentioned earlier when introducing the kernel deconvolution estimator.

The densities were estimated using the sinc kernel, $\frac{sin(x)}{\pi x}$. The sinc kernel has attractive properties for deconvolution, namely a Fourier transform that is is flat around zero and which has bounded support, the later of which provides integrability. Other kernels which omit a flat Fourier transform near zero were tried (triangular and flat), but each produced results which oscillated wildly.

Table 8 provide quantiles for the three distributions. While the quantiles between the nonparametric and the LaPlace approaches differ somewhat, all three approaches give somewhat similar results.

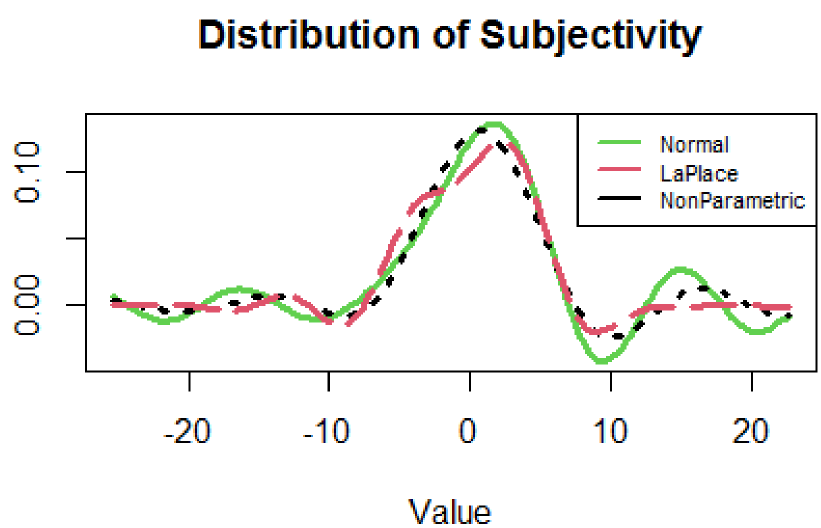
[8]Note that the use of a Normal or LaPlace parametrization for f_ϵ does not force $f_{\widetilde{S}}$ to have a Normal or LaPlace form. This can be seen by the fact that the nonparametric estimate of f_ϵ produces a similar shape.

Figure 5: Residuals from Azic and Lame



This plots densities based on the residuals in Azic and Lamé (2018). NoInfo corresponds to the treatment where managers were not provided with subjective statements, while Info corresponds to the treatment where managers did receive subjective statements.

Figure 6: Distribution of Subjectivity



This figure plots the estimated kernel deconvolution densities based on the estimator. The non-parametric density is done using the sinc kernel to estimate f_ϵ. The Normal and LaPlace densities are estimated by first fitting the error sample to a Normal and LaPlace distribution, respectively, via maximum likelihood, and then using deconvolution to recover the density estimate. Bandwidth selection is chosen via cross-validation. Estimation is done via the *kerdec* R package (see [8]).

Table 8: Quantiles for the Estimated f_S

Quantile	**Normal** f_ϵ	**NonParametric** f_ϵ	**LaPlace** f_ϵ
Median	0.43	0.33	0.48
10	-4.04	-3.49	-3.99
25	-1.84	-1.75	-2.16
75	2.29	2.28	2.61
90	3.47	3.68	3.91

Note: The table above provides quantiles for the distribution of f_S given the distributions estimated in Figure 6. The columns measure quantiles for the CDF corresponding to f_S under the settings where $f_\epsilon$is (i) fitted to a normal; (ii) estimated non-parametrically; and (iii) fitted to a LaPlace.

To put the results in Table 8 in context, it is necessary to consider them alongside the statement quality as judged by the proxy in [6] (i.e. the $\bar{P}_i$ in the model section of this paper). In [6], proxy quality scores of statements ranged from 10.08 to 21.9, with a median quality score of 14.58, a mean quality score of 14.96, and a standard deviation of 2.72.

Then, for a worker who's statement is judged to be of median proxy-quality (scored a 14.58), there is roughly a 25% chance that his statement quality will be judged to be below 12.58 by a given manager, and a roughly 25% chance that his statement quality will be judged to be above a 16.8. In other words, there is a 50% chance that a manager will interpret the median worker's score to be between (roughly) 86% and 115% of the quality described by the proxy measure; the remaining 50% of the time, the disagreement between the quality proxy and a manager's view of statement quality is even greater.

A natural extension for the results above to see what makes a statement more or less subjective. The procedure in this paper recovers the distribution of a latent variable, not its realized values. Therefore, no regression of subjectiveness on worker, manager, or statement characteristics can be run. Instead, the original sample has to be partitioned into sub-samples, from which f_S can be estimated and compared. Unfortunately, several attempts to do this (e.g. splitting workers into male/female, splitting statements by length) were unsuccessful; results were extremely jumpy – especially when f_ϵ was estimated nonparametrically – and the level of disagreement between the three densities seen in Figure 1, was extremely large. This is likely due to the estimator's poor convergence properties and the smaller sample sizes.

Bibliography

[1] John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

[2] Oktay J. Akkus, Anthony Cookson, and Ali Hortacsu. The determinants of bank mergers: A revealed preference approach. *Management Science*, 62(8):2241–2258, 2015.

[3] Joseph G. Altonji, Lisa B. Kahn, and Jamin D. Speer. Cashier or consultant? entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, 34(S1):S361–S401, 2016.

[4] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[5] Mark Azic. Subjectivity in hiring: Theory with evidence from professional baseball contracts. *Working Paper*, 2019.

[6] Mark Azic and Diego Lamé. The effect of subjective information in hiring decisions. *Working Paper*, 2018.

[7] George Baker, Robert Gibbons, and Kevin J. Murphy. Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics*, 11(1):1125–1156, 1994.

[8] Guillermo Basulto-Elias. Kernel deconvolution density estimation. *Doctoral dissertation, Iowa State University, Ames, Iowa*, 2016.

[9] Joseph Berkson. Are there two regressions? *Journal of the American Statistical Association*, 45(250):164–180, 1950.

[10] Iris Bohnet, Alexandra Van Geen, and Max Bazerman. When performance trumps gender bias: joint versus separate evaluation. *Management Science*, 62(5):1225–1234, 2015.

[11] Jasmijn C. Bol. Subjectivity in compensation contracting. *Journal of Accounting Literature*, 27:1–24, 2008.

[12] Jasmijn C. Bol. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5):1549–1575, 2011.

[13] Kenneth Burdett and Dale T Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273, 1998.

[14] David Ruppert Leonard A. Stefanski Carroll, Raymond J and Ciprian M. Cariniceanu. *Measurement Error in Nonlinear Models*. Chapman Hall/CRC, 2006.

[15] Raymond J. Carroll and Peter Hall. Optimal rates of convergence for deconvoluting a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.

[16] Pierre-Andre Chiappori and Bernard Salanie. The econometrics of matching models. *Journal of Econommic Literature*, 54(3):832–861, 2016.

[17] Katherine B. Coffman, Christine L. Exley, and Muriel Niederle. When gender discrimination is not about gender. *Harvard Business School Working Paper*, 2017.

[18] David J. Deming. Growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640, 2017.

[19] Peter J. Diggle and Peter Hall. A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society*, 55(2):523–531, 1993.

[20] Henry S Farber. The analysis of interfirm worker mobility. *Journal of Labor Economics*, 12(4):554–593, 1994.

[21] Jeremy T Fox. Semiparametric estimation of multinomial discrete choice models using a subset of choices. *RAND Journal of Economics*, 38(4):1002–1019, 2007.

[22] Jeremy T Fox. Identification in matching games. *Quantitative Economics*, 1(2):203–254, 2010.

[23] Jeremy T. Fox. Estimating matching games with transfers. *Quantitative Economics*, 9(1):1–38, 2018.

[24] Jeremy T. Fox and Patrick Bajari. Measuring the efficiency of an fcc spectrum auction. *American Economic Journal: Microeconomics*, 5(1):100–146, 2013.

[25] Jeremy T. Fox, Chenyu Yang, and David Hsu. Unobserved heterogeneity in matching games. *Jounral of Political Economy*, 126(4):1339–1371, 2018.

[26] David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[27] David Gale and Lloyd S Shapley. Job matching with heterogeneous firms and workers. *Econometrica*, 49(2):437–450, 1981.

[28] Vince Gennaro. *Diamond Dollars: The Economics of Winning in Baseball*. Diamond Analytics, 2013.

[29] Matthew Gentry, Tatiana Komarova, and Pasquale Schiraldi. Preferences and performance in simultaneous first-price auctions: A structural analysis. *Working Paper*, 2016.

[30] Claudia Goldin and Cecilia Rouse. Orchestrating impartiality: the impact of blind auditions on female musicians. *American Economic Review*, 90(4):715–741, 2000.

[31] Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000.

[32] Milton Harris and Artur Raviv. Differences of opinion make a horse race. *The Review of Financial Studies*, 6(3):473–506, 1993.

[33] Mitchell Hoffman, Lisa B. Kahn, and Danielle Li. Discretion in hiring. *The Quarterly Journal of Economics*, 133(2):765–800, 2017.

[34] Joel L. Horowitz and Marianthi Markatou. Semiparametric estimation of regression models for panel data. *Review of Economic Studies*, 63:145–168, 1996.

[35] Boyan Jovanovic. Job matching and theory of turnover. *Journal of Political Economy*, 87(5):972–990, 1979.

[36] Chinhui Juhn, Kevin M. Murphy, and Brooks Pierce. Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–442, 1993.

[37] Lisa B. Kahn. The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316, 2010.

[38] Eugene Kandel and Neal D. Pearson. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4):831–872, 1995.

[39] Alexander S. Jr Kelso and Vincent P. Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica*, 50(6):1483–1504, 1982.

[40] Yunmi Kong. Sequential auctions with synergy and affiliation across auctions. *Journal of Political Economy, forthcoming.*

[41] Alan Krueger and Lawrence Summers. Efficiency wages and the inter-industry wage structure. *Econometrica*, 56(2):259–293, 1988.

[42] Edward P. Lazear and Kathryn L. Shaw. Introduction: Firms and distribution of income: The roles of productivity and luck. *Journal of Labor Economics*, 36(S1):S1–S12, 2018.

[43] Michael Lewis. *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton Company, 2004.

[44] Qi Li, Esfandiar Maasoumi, and Jeffrey S Racine. A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, 148(2):186–200, 2009.

[45] Jeremy Lise, Costas Meghir, and Jean-Marc Robin. Matching, sorting, and wages. *Review of Economic Dynamics*, 19(1):63–87, 2016.

[46] Hamish Low, Costas Meghir, and Luigi Pistaferri. Wage risk and employment risk over the life cycle. *American Economic Review*, 100(4):1432–1467, 2010.

[47] Esfandiar Maasoumi and Jeffrey S Racine. Entroy and predictability of stock market returns. *Journal of Econometrics*, 107:291–312, 2002.

[48] Charles F Manski. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3:205–228, 2000.

[49] John McMillan. *Reinventing the Bazaar: A Natural History of Markets*. Norton & Company, 2002.

[50] Paul Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, 2004.

[51] Paul Milgrom and Nancy Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27, 1982.

[52] Frank Moers. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations, and Society*, 30(1):67–80, 2005.

[53] Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

[54] Philip Oreopoulos, Till Von Wachter, and Andrew Heisz. The short and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29, 2012.

[55] Paul Oyer and Scott Schaefer. Personnel economics: Hiring and incentives. *Handbook of Labor Economics*, 4B, 2011.

[56] Richard Paulsen. Three essays in productivity and earnings. *Doctoral dissertation, Northeastern University, Boston, Massachusetts*, 2019.

[57] Lewis J Pollis. If you build it: Rethinking the market for major league baseball front office personnel. *Senior Honors Thesis, Brown University, Providence, Rhode Island*, 2014.

[58] Fabien Postel-Vinay and Jean-Marc Robin. Equilibrium wage dispersion with worker and firm heterogeneity. *Econometrica*, 70(6):2295–2350, 2002.

[59] Ernesto Reuben, Paola Sapienza, and Luigi Zingales. How stereotypes impair women careers in science. *Proceedings of the National Academy of Sciences*, 201314788., 2014.

[60] Alvin E Roth and Marilda A. Oliveira Sotomayor. *Two-sided matching: A Study in Game-Theoretic Modeling and Analysis.* Cambridge University Press, 1990.

[61] Ariel Rubinstein. On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy*, 101(3):473–484, 1993.

[62] Alan Schwarz. *The Numbers Game: Baseball's Lifelong Fascination with Statistics*. St. Martin's Griffin, 2005.

[63] Leonard A Stefanski and Raymond J Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2):169–184, 1990.

[64] Robert H. Topel and Michael P. Ward. Job mobility and the careers of young men. *Quarterly Journal of Economics*, 107(2):439–479, 1992.

[65] Fischbacher Urs. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.

[66] Domonkos F Vamossy. Investor emotions and earnings announcements. *Working Paper*, 2020.

[67] William Witnauer and Richard Rogers. Major league baseball career length in the 20th century. *Population Research and Policy Review*, 26(4):371–386, 2007.