Two-Tiered Stochastic Frontier Models: A Bayesian Perspective

SHIRONG ZHAO* JEREMY LOSAK†

August 2022

Abstract

Bayesian methods have been well studied in one-tiered stochastic frontier literature, but as of yet have not been proposed in a two-tiered stochastic frontier (2TSF) setting. Recently, 2TSF models have drawn much attention, observed by increased theoretical extensions and empirical applications. This paper fills the gap in the literature by presenting a Bayesian approach to estimating 2TSF models, with and without efficiency (or bargaining power) determinants. Posterior distributions for the parameters and efficiencies for two bargaining parties are derived. Finally, we use both maximum likelihood estimation and Bayesian methods to analyze bargaining power in Major League Baseball salary arbitration negotiations. We find that players who generate their value by hitting for power have more bargaining power, while players who generate their value by having high on-base abilities have less bargaining power. This result is consistent with pre-Moneyball free agent market valuations for these skills.

Keywords: Two-Tiered Stochastic Frontier, Bayesian Econometrics, Gibbs Sampling, Moneyball, Final Offer Arbitration

JEL Classification: C1, C3, Z22

Acknowledgement

We appreciate comments from Christopher Parmeter when we started this project. We also thank Shannon Losak for her careful edits. Shirong Zhao appreciates the financial support from Dongbei University of Finance and Economics. All remaining errors are our own.

Declaration of Competing interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence this paper's results.

*Zhao: School of Finance, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China; email: shironz@dufe.edu.cn; ORCiD: 0000-0002-5306-7685

†Losak (Corresponding Author): David B. Falk College of Sport and Human Dynamics, Syracuse University, Syracuse, NY 13244, United States; email: jmlosak@syr.edu; ORCiD: 0000-0002-4591-2762

1 Introduction

Compared to the typical one-tiered stochastic frontier (SF) (Aigner et al. 1977; Meeusen and van den Broeck 1977), the two-tiered stochastic frontier (2TSF) (Polachek and Yoon 1987) is more complicated. While the one-tiered SF has a composite error consisting of a noise term and a non-negative efficiency term, the 2TSF composite error term has three components rather than two. More specifically, the 2TSF has a typical noise term, and two non-negative efficiency terms with opposite signs. Unlike the one-tiered SF, where the non-negative efficiency term is typically used to measure a firm's efficiency, the two non-negative efficiency terms in 2TSF are typically used to measure the bargaining powers or asymmetric information of two different parties, where one party has a positive effect on the outcomes while the other has a negative effect.

Due to the embedded econometric model structure of 2TSF where there are two non-negative terms that have opposite signs, it is natural to employ this model to study economic outcomes (e.g., the prices paid/received) in various markets, especially those that have two bargaining parties (e.g., the buyers and sellers). The 2TSF can be used to measure the strength of the bargaining power of each party, the surplus extracted by each party, or the incomplete information in the market. Given their obvious applicability to many economic questions, 2TSF models have recently drawn much attention, observed by increased theoretical extensions (Polachek and Yoon 1987; Groot and Oosterbeek 1994; Polachek and Yoon 1996; Kumbhakar and Parmeter 2010; Papadopoulos 2015; Wang 2017; Das and Polachek 2017; Parmeter 2018; Papadopoulos et al. 2021; Papadopoulos 2021b; Papadopoulos 2022; etc) and has also been applied in a number of empirical settings (Polachek and Yoon 1987; Ferona and Tsionas 2012; Kumbhakar and Parmeter 2010; Xu et al. 2021; Hu and Pei 2020; Papadopoulos 2021a; etc). See Papadopoulos (2018) for a comprehensive review on the methods and applications of 2TSF models.

Several methods have been developed in the literature to estimate one-tiered SF models, including maximum likelihood estimation (MLE), corrected ordinary least squares (COLS), Bayesian methods (Van den Broeck et al. 1994; Greene 2005), and Quantile methods (Jradi et al. 2019; Tsionas 2020; Tsionas et al. 2020; Jradi et al. 2021; Zhao 2021; Papadopoulos and Parmeter 2022). However, for 2TSF models, maximum likelihood estimation (MLE) is the predominant method. To this point, a Bayesian approach has not yet been formally introduced, a clear gap in the literature.

Since their introduction by Van den Broeck et al. (1994), Bayesian analyses are widely used in the one-tiered SF literature. Some early influential papers include Kim and Schmidt (2000), Koop et al. (2001), Griffin and Steel (2004), Greene (2005), Griffin and Steel (2007), etc. Some recent Bayesian papers in the one-tiered SF include Tsionas and Mallick (2019), Klein et al. (2020), Kumbhakar and Tsionas (2021), and Tsionas and Kumbhakar (2021).

In this paper, we fill a gap in the literature by extending the Bayesian methods for one-tiered SF models (Van den Broeck et al. 1994; Greene 2005; etc) to 2TSF settings. Although the Bayesian methods for 2TSF models seem to be straightforward, to the best of our knowledge, we are the first to formally formulate the Bayesian framework and thus provide an alternative estimation technique for 2TSF models. Further, similar to the one-tiered SF models, we incorporate two different situations—without and with efficiency/bargaining power determinants (or environmental variables that could impact individual or firm bargaining power) (Kumbhakar et al. 1991; Reifschneider and Stevenson 1991).

Due to the complex structure of 2TSF, its MLE estimation typically involves optimization methods and numerical integration methods. However, the Markov Chain Monte Carlo (MCMC) in the Bayesian methods is naturally appropriate for the 2TSF settings. In the context of 2TSF models, the benefit of using a Bayesian approach is that we account for the uncertainty of the parameter when deriving the posterior distributions, which are straightforward to derive. Moreover, the Bayesian approach does not involve the numerical complexities. The typical tradeoff of using a Bayesian approach is that it usually requires a lot of computation power; however, this is less of an issue with rapid growth in technological and computational resources.

Finally, we illustrate our proposed method using an empirical application: bargaining power in Major League Baseball (MLB) arbitration salary negotiations. This market was previously explored using a deterministic double frontier approach by Hadley and Ruggiero (2006) focused on race and bargaining power. In the empirical example, we utilize our new Bayesian methods to consider relative bargaining power as a function of specific player characteristics. Conditional on a player's production levels, do certain player characteristics impact bargaining power and influence final awarded arbitration salary? Sport labor market data provide a great empirical laboratory to test economic theories around worker pay and performance, especially with the abundance of team performance, worker productivity, and salary data—all measurable and publicly available for free (Kahn 2000). Given the two bargaining parties (player and team), realization of both inputs (player production) and outputs (player salary) of the arbitration process, and potential environmental variables impacting bargaining position (player characteristics), baseball arbitration provides the perfect case study to introduce our new Bayesian techniques and compare to MLE.

The rest of the paper is organized as follows. Section 2 presents a brief literature review related to 2TSF framework. Section 3 introduces our Bayesian approach for 2TSF without and with inefficiency/bargaining power determinants. Section 4 uses Major League Baseball salary arbitration negotiations to illustrate our Bayesian approach, and we also compare the results of Bayesian methods and MLE. Section 5 concludes and recommends future research directions.

2 Literature Review

One-tired SF is related to 2TSF, and it has been well studied in the literature. Readers can refer to Kumbhakar et al. (2020a) and Kumbhakar et al. (2020b) for a brief review. In this section, we focus our review on two-tiered stochastic frontier models. We first review the literature related to the theoretical developments of 2TSF, then briefly mention various markets that 2TSF have been applied to in the literature. Finally, we review papers relevant to our empirical example of Major League Baseball salary arbitration negotiations.

2.1 Literature Related to 2TSF

Polachek and Yoon (1987) first introduced 2TSF models, which were used to estimate earnings functions in the labor market. The two non-negative terms in Polachek and Yoon (1987) represent the ignorance of employee and employer due to incomplete information. That is, the employee does not know all the possible work offers, and thus receives less than the highest available offer. On the other hand, the employer does not know the reservation wage of all potential employees, and thus pays more than the lowest possible reservation wage. As an extension to the basic 2TSF models in Polachek and Yoon (1987), Polachek and Yoon

(1996) propose panel estimates of 2TSF models while simultaneously accounting for heterogeneity. By applying this new model to estimate an earnings function, Polachek and Yoon (1996) found that employers typically have more information than employees, giving them more bargaining power in salary negotiations. Das and Polachek (2017) proposes a novel 2TSF by embedding heterogeneity into the composite error term to estimate labor market joiners and leavers. This novel 2TSF is identified using a two-step estimation approach.

Note that both Polachek and Yoon (1987) and Polachek and Yoon (1996) assume that the noise term follows the Normal distribution while the two non-negative terms (or efficiency terms) follow Exponential distributions. However, it is prevailing to assume the efficiency term to follow Half-Normal distribution in SF models (or one-tiered models). Papadopoulos (2015) then proposes 2TSF with Half-Normal specifications, where the two non-negative terms are assumed to follow Half-Normal distributions. Papadopoulos (2021b) proposes 2TSF with Generalized Exponential specifications. There are also other specifications, such as the Truncated-Normal specification (Wang 2017), which nests the Half-Normal.

Groot and Oosterbeek (1994) extend the 2TSF by incorporating efficiency determinants (i.e., the observable characteristics that could affect bargaining power) into the model by assuming a linear function between the mean value of efficiency/bargaining power and the observable characteristics. This approach is problematic as the efficiency/bargaining power is non-negative. By following the standard practice in the one-tiered SF, Kumbhakar and Parmeter (2010) models the parameters in the Exponential distributions as exponential functions of observable characteristics. Parmeter (2018) further extends the scaling property that is well-studied in SF literature (this property is used to measure the effects of efficiency determinants) to the 2TSF models, which could avoid the restrictive parametric distributional assumptions (i.e., Exponential or Half-Normal distributions) on the efficiency terms. Papadopoulos et al. (2021) propose three different approaches to model the dependence of the bargaining power in the 2TSF models. More recently, Papadopoulos (2022) provides a different view on 2TSF from a bilateral Nash bargaining model under value uncertainty and asymmetric information. Papadopoulos (2022) also uses Copulas to take endogeneity into account.

Besides the above theoretical developments of the 2TSF, there has been an increasing number of empirical applications of 2TSF, including those in the labor market (Polachek and Yoon 1987; Groot and Oosterbeek 1994; Polachek and Yoon 1996; Sharif and Dar 2007; Murphy and Strobl 2008; Kumbhakar and Parmeter 2009; Blanco 2017; Das and Polachek 2017), housing market (Kumbhakar and Parmeter 2010; Pu et al. 2022), energy market (Liu et al. 2019), health market (Tomini et al. 2012; Gaynor and Polachek 1994; Chawla 2002), water rights market (Xu et al. 2021), auctions market (Ferona and Tsionas 2012; Verteramo Chiu et al. 2022), wine market (Fried and Tauer 2019), and other related markets (Wang 2018; Zhang et al. 2018; Hu and Pei 2020; Papadopoulos 2021a; etc).

2.2 Literature Related to Major League Baseball Salary Arbitration Negotiations

To the best of our knowledge, stochastic 2TSF models have not yet been applied in Major League Baseball salary arbitration negotiations, which is a natural market that involves bargaining between players and teams. Hadley and Ruggiero (2006) utilize a deterministic double frontier model to estimate, among other things, the impact of race on bargaining power in baseball arbitration salary outcomes. In their approach, they first calculate an upper player frontier (most a player can expect to earn) and lower team frontier (minimum a team can expect to spend). Next, they calculate relative efficiency estimates for players in their sample. Finally, they regress these efficiency estimates on various characteristics, including variables controlling for a player's race. While this empirical strategy was certainly novel, it lacks theoretical support in the literature. Conversely, 2TSF models have been used in many empirical applications of labor market bargaining, power as previously discussed.

After introducing our Bayesian approaches of 2TSF, we use data from Major League Baseball salary arbitration negotiations over 2000–2017 to illustrate our Bayesian estimation methods. Peculiarities of the baseball labor market, including baseball arbitration, are well documented in the sport economic literature. Scully (1974) is the first to estimate the marginal revenue product and degree of exploitation of Major League Baseball players. A critical area of interest is baseball labor market efficiency, especially before and after the time of the groundbreaking book Moneyball (Hakes and Sauer 2006; Hakes and Sauer 2007; Brown et al. 2017; Holmes et al. 2018; Pinheiro and Szymanski 2022). This line of research considers if various player production attributes are efficiently priced into player salaries, specifically the returns to getting on base and returns to hitting for power. We consider whether these same hitter attributes impact bargaining power in baseball arbitration cases. This adds to the collective literature on baseball arbitration (Wittman 1986; Dworkin 1977; Faurot and McAllister 1992; Gustafson and Hadley 1995; Fizel 1996; Miller 2000; Faurot 2001; Fizel et al. 2002; Farmer et al. 2004; Hadley and Ruggiero 2006 Hanany et al. 2007).

3 The Two-Tiered Stochastic Frontier Model

3.1 The 2TSF Model Without Efficiency Determinants

The 2TSF is specified by,

$$y_i = x_i\beta + v_i - u_i + w_i = x_i\beta + \varepsilon_i, \tag{3.1}$$

where y_i is the explained variable; x_i is the $1 \times p$ explanatory vector; v_i is the symmetric noise term; u_i and w_i are non-negative efficiency measurements from two parties; and $\varepsilon_i = v_i - u_i + w_i$ is the composite error term. The 2TSF has a three-component error term of v_i, u_i and w_i, compared to the SF which has a two-component error term. The SF has either $v_i - u_i$ for a production frontier, or $v_i + w_i$ for a cost frontier. The equilibrium value for y_i is measured by $x_i\beta$. For 2TSF, bargaining power is introduced, with u_i for one party (a "buyer", for example) and w_i for the other (a "seller", for example). These parties have the opposite effects on the explained variable y_i. It is standard to assume v_i, u_i and w_i are independent of each other, and independent from x_i, for all bargains i.

For future reference, we denote $n \times 1$ vector $y = (y_1, y_2, \ldots, y_n)'$, $v = (v_1, v_2, \ldots, v_n)'$, $u = (u_1, u_2, \ldots, u_n)'$, $w = (w_1, w_2, \ldots, w_n)'$, and $n \times p$ matrix $x = (x_1, x_2, \ldots, x_n)'$. Further, $N(a, b)$ denotes the univariate Normal distribution with mean a and variance b. $N_p(c, d)$ denotes the p-variate Normal distribution with mean c and co-variance d. $\mathcal{E}(\lambda)$ denotes the Exponential distribution with rate λ. $\mathcal{G}(e, f)$ denotes the Gamma distribution with shape e and rate f. $N_+(0, \sigma^2)$ denotes the Half-Normal distribution. $I(g)$ is the indicator function.

In this paper, we focus on 2TSF with Exponential specifications as it is predominant in the literature. In our appendix, we also derive the corresponding Bayesian results for 2TSF with Half-Normal specifications for researchers that are interested in this specification. The Exponential 2TSF has the Normal-Exponential-Exponential distribution, i.e., $v_i \sim N(0, \sigma_v^2)$, $u_i \sim \mathcal{E}(\lambda_1^{-1})$, and $w_i \sim \mathcal{E}(\lambda_2^{-1})$. We denote $\phi = 1/\sigma_v^2$, which will be shown to be convenient in deriving the conditional posterior distributions for ϕ. The conditional density for y_i has a closed form (Kumbhakar and Parmeter 2009) and maximum likelihood estimation can be utilized to recove the parameters.

To complete the Bayesian model, we must specify priors for the parameters. Following the SF literature, we assume that,[1]

$$\beta \propto \text{cons}, \phi \sim \mathcal{G}(a_0, b_0), \lambda_1^{-1} \sim \mathcal{G}(a_1, b_1), \lambda_2^{-1} \sim \mathcal{G}(a_2, b_2). \tag{3.2}$$

With the prior specification, we have the complete conditional posterior distributions,

$$\beta|y, x, u, w, \phi \sim N_p\left((x'x)^{-1}x'(y+u-w), (x'x)^{-1}\phi^{-1}\right), \tag{3.3}$$

$$\phi|y, x, u, w, \beta \sim \mathcal{G}\left(a_0 + \frac{n}{2}, b_0 + \frac{1}{2}\sum_{i=1}^{n}(y_i + u_i - w_i - x_i\beta)^2\right), \tag{3.4}$$

$$\lambda_1^{-1}|u \sim \mathcal{G}\left(a_1 + n, b_1 + \sum_{i=1}^{n} u_i\right), \tag{3.5}$$

$$\lambda_2^{-1}|w \sim \mathcal{G}\left(a_2 + n, b_2 + \sum_{i=1}^{n} w_i\right), \tag{3.6}$$

$$u_i|y, x, w, \beta, \phi, \lambda_1^{-1} \sim N\left(x_i\beta + w_i - y_i - \lambda_1^{-1}\phi^{-1}, \phi^{-1}\right) I(u_i \geq 0), \tag{3.7}$$

and

$$w_i|y, x, u, \beta, \phi, \lambda_2^{-1} \sim N\left(y_i + u_i - x_i\beta - \lambda_2^{-1}\phi^{-1}, \phi^{-1}\right) I(w_i \geq 0). \tag{3.8}$$

The Gibbs sampling method (Geman and Geman 1984) can be used to sequentially draw random numbers from the above conditional posterior distributions. Inferences can be made based on the posterior random values for the parameters. With these conditional posterior distributions, the following quantities can be estimated (surplus extracted by each party as well as net surplus): $E(u_i)$, $E(w_i)$, $E(u_i - w_i)$, $E(e^{-u_i})$, $E(e^{-w_i})$, and $E(e^{-u_i} - e^{-w_i})$. Note $E(u_i)$ and $E(w_i)$ measure the surplus (in absolute value) extracted by buyer and seller, while $E(e^{-u_i})$ and $E(e^{-w_i})$ measure the extracted surplus in percentage value, and $E(u_i - w_i)$ and $E(e^{-u_i} - e^{-w_i})$ measure the net surplus. See Kumbhakar and Parmeter (2009) for more details.

[1] Instead of a flat prior, we can assume β follows a p-variate Normal distribution. Further, we can impose economic regularity constraints for β.

Note that in the Bayesian framework, we could use the posterior draws to directly calculate the above quantities. However, for MLE we need to rely on the Jondrow et al. (1982) measures, which decompose the composite error into inefficiency and noise components through using the conditional measures (conditional on the composite error ε_i). The corresponding measures for 2TSF models are derived in equations (9), (10), (A.10), and (A.13) of Kumbhakar and Parmeter (2009) for $E(u_i \mid \varepsilon_i)$, $E(w_i \mid \varepsilon_i)$, $E(e^{-u_i} \mid \varepsilon_i)$ and $E(e^{-w_i} \mid \varepsilon_i)$, respectively.[2]

3.2 The 2TSF Model With Efficiency Determinants

Researchers sometimes might want to take into account observable characteristics that could impact the bargaining power/incomplete information of two opposed parties. Following Kumbhakar and Parmeter (2010), we model the parameters in the Exponential distribution as functions of observable characteristics.

To include variables that affect u_i and w_i, we assume $u_i \sim \mathcal{E}(\lambda_{1i}^{-1})$ and $w_i \sim \mathcal{E}(\lambda_{2i}^{-1})$, where $\lambda_{1i} = \exp(z_{1i}\delta_1)$, $\lambda_{2i} = \exp(z_{2i}\delta_2)$, and z_{1i} and z_{2i} are $1 \times q$ and $1 \times r$ vectors, respectively, of efficiency determinants. The effects of z_{1i} on u_i are achieved through affecting its mean value λ_{1i}, while the effects of z_{2i} on w_i are achieved through λ_{2i}. This approach to estimating efficiency determinants is standard in one-tiered SF literature, see Kumbhakar et al. (1991) and Reifschneider and Stevenson (1991); they have also been discussed or studied in 2TSF by Kumbhakar and Parmeter (2010), but in the MLE framework. Formally, we have the following 2TSF model with efficiency determinants, where the effects of efficiency determinants are specified in a latent class,

$$\begin{aligned} y_i =& x_i\beta + v_i - u_i + w_i, \\ & u_i \sim \mathcal{E}(\lambda_{1i}^{-1}), \lambda_{1i} = \exp(z_{1i}\delta_1), \\ & w_i \sim \mathcal{E}(\lambda_{2i}^{-1}), \lambda_{2i} = \exp(z_{2i}\delta_2), \\ & v_i \sim N(0, 1/\phi), \end{aligned} \tag{3.9}$$

with the following prior for the parameters,

$$\beta \propto \text{cons}, \phi \sim \mathcal{G}(a_0, b_0), \delta_1 \sim N_q(0, R_1), \delta_2 \sim N_r(0, R_2). \tag{3.10}$$

[2] As noted by Papadopoulos (2018), there are typo errors in equations (11) and (12) of Kumbhakar and Parmeter (2009). Thus, researchers should use equations (A.10) and (A.13) instead of equations (11) and (12).

Conditional posterior distributions for β and ϕ are the same as the case without efficiency determinants and are given in (3.3) and (3.4), while those for the remaining parameters are,

$$\delta_1|z_1, u \sim \prod_{i=1}^{n} \Big(\frac{1}{\exp(z_{1i}\delta_1)} \exp(-\frac{u_i}{\exp(z_{1i}\delta_1)}) \Big) \exp(-\frac{\delta_1' R_1^{-1} \delta_1}{2}), \tag{3.11}$$

$$\delta_2|z_2, w \sim \prod_{i=1}^{n} \Big(\frac{1}{\exp(z_{2i}\delta_2)} \exp(-\frac{w_i}{\exp(z_{2i}\delta_2)}) \Big) \exp(-\frac{\delta_2' R_2^{-1} \delta_2}{2}), \tag{3.12}$$

$$u_i|y, x, z_1, w, \beta, \phi, \delta_1 \sim N \left(x_i\beta + w_i - y_i - \frac{1}{\exp(z_{1i}\delta_1)} \phi^{-1}, \phi^{-1} \right) I(u_i \geq 0), \tag{3.13}$$

and

$$w_i|y, x, z_2, u, \beta, \phi, \delta_2 \sim N \left(y_i + u_i - x_i\beta - \frac{1}{\exp(z_{2i}\delta_2)} \phi^{-1}, \phi^{-1} \right) I(w_i \geq 0). \tag{3.14}$$

Note that we have complete conditional posterior distributions for β, ϕ, u, and w, and hence the Gibbs sampling method can be used. However, for δ_1 and δ_2, we cannot derive complete conditional posterior distributions; instead, we use Metropolis' algorithm (Metropolis et al. 1953).

It is possible for efficiency determinants to exist for w_i, but not for u_i. Holding everything the same as above, here, we assume $u_i \sim \mathcal{E}(\lambda_1^{-1})$. Further, assume the prior for $\lambda_1^{-1} \sim \mathcal{G}(a_1, b_1)$. The conditional posterior distributions for β, ϕ, δ_2, and w_i are unchanged and are listed in (3.3), (3.4), (3.12), and (3.14), respectively. Conditional posterior distributions for λ_1^{-1} and u_i are

$$\lambda_1^{-1}|u \sim \mathcal{G} \left(a_1 + n, b_1 + \sum_{i=1}^{n} u_i \right), \tag{3.15}$$

and

$$u_i|y, x, w, \beta, \phi, \lambda_1^{-1} \sim N \left(x_i\beta + w_i - y_i - \lambda_1^{-1}\phi^{-1}, \phi^{-1} \right) I(u_i \geq 0), \tag{3.16}$$

respectively.

On the other hand, it is possible that there are only efficiency determinants for u_i, but not for w_i. For this situation, we assume $w_i \sim \mathcal{E}(\lambda_2^{-1})$. Further, assume the prior for $\lambda_2^{-1} \sim \mathcal{G}(a_2, b_2)$. The conditional posterior distributions for β, ϕ, δ_1 and u_i are unchanged and are listed in (3.3), (3.4), (3.11) and (3.13). Conditional posterior distributions for λ_2^{-1} and w_i are

$$\lambda_2^{-1}|w \sim \mathcal{G} \left(a_2 + n, b_2 + \sum_{i=1}^{n} w_i \right), \tag{3.17}$$

and

$$w_i|y, x, u, \beta, \phi, \lambda_2^{-1} \sim N\left(y_i + u_i - x_i\beta - \lambda_2^{-1}\phi^{-1}, \phi^{-1}\right) I(w_i \geq 0), \tag{3.18}$$

respectively.

4 An Empirical Illustration

We illustrate our Bayesian 2TSF method with efficiency determinants by using Major League Baseball (MLB) arbitration data to identify player characteristics that impact bargaining position in salary negotiations. If a player has between three and six years of Major League service, he can elect for final-offer salary arbitration (FOA).[3] As part of the FOA process, the team and player submit salary figures to an arbitration panel. The panel, consisting of three randomly selected (from a pre-approved pool) non-baseball expert labor arbitrators, chooses which figure to award as the player's salary. They make their determinations based on the following criteria: the quality of the player's contributions during the past season, the length and consistency of career contributions, player past compensation, comparative baseball salaries, the existence of player physical or mental defects, and contributions to the club's recent performance. In a hearing, the panel selects one of the two submitted figures and cannot pick a middle-ground number. This setup incentivizes parties to settle prior to reaching the panel, due to exposed risk both sides face in a hearing—the difference between submissions are often worth hundreds of thousands (if not millions) of dollars. Preparing a case for an arbitration panel also requires significant labor hours and resources from both the player's and team's representatives, further motivating a settlement outcome. In fact, from 2000–2017, the vast majority (about 96%) of all eligible cases resulted in settlements.

Baseball arbitration has been examined previously in the literature. Dworkin (1977) and Wittman (1986) model the decision process of the arbitration panel. They postulate that the panel determines a "true" arbitration value and then selects the submitted salary figure closest to that value. Of course, the two parties know this and submit salary figures that will maximize (minimize) expected player (team) salary (cost). Numerous papers focus on the bargaining/negotiation process, typically using a Nash Bargaining framework (Dworkin

[3] Players with less than three years of service time could also be eligible to file for arbitration if they rank within the top 22% of service time for players with between two and three years of service. These players are known as Super Two players.

1977; Faurot and McAllister 1992). Hanany et al. (2007) emphasize the importance of relative risk preferences in bargaining outcomes. It illustrates that depending on the relative risk preferences of the two parties, there exists a set of settlement outcomes that dominates the expected utility from going to a hearing. Player risk preferences are also considered in Faurot (2001), among others.

4.1 Empirical Approach

Players seek to maximize salary, while teams seek to minimize payroll expenditures.[4] In this illustration, we analyze how player production attributes impact bargaining position during settlement negotiations. Arbitrators may evaluate similarly valuable players differently based on how they produce value. For example, a player who predominantly hits home runs may be compensated differently than a player who prioritizes getting on-base, even if their aggregate net value to the team, as perceived by the arbitration panel, is the same.

Formally, we estimate the following 2TSF using MLB arbitration-eligible cases from 2000–2017 (723 observations),

$$\begin{aligned} \log(y_i) =& x_i\beta + v_i - u_i + w_i, \\ & v_i \sim N(0, 1/\phi), \\ & u_i \sim \mathcal{E}(\lambda_1^{-1}), \\ & w_i \sim \mathcal{E}(\lambda_{2i}^{-1}), \lambda_{2i} = \exp(z_{2i}\delta_2), \end{aligned} \tag{4.1}$$

where y_i is an output, x_i are explanatory variables, and z_{2i} are player characteristics that influence player bargaining power (w_i). We consider two outputs for y_i: salary and raise.[5] Our official salary details come from a proprietary source, although much of our data is also publicly available (especially recent data). In addition to the player's awarded salary, we analyze a player's salary raise over the previous season since a player's "previous compensation" is one factor included in the arbitration criteria.

[4] A team's true objective function may be to maximize profits or wins. Acquiring players via free agency (win maximizing) usually comes at a high expense (and may not be profit maximizing). However, a player going through arbitration is bound to the team he is negotiating with unless the team chooses to release the player. Therefore, profit maximization is achieved in this setting by limiting what the player earns in salary arbitration. Win maximization is not impacted by the salary arbitration process.

[5] All salary and raise figures are adjusted using GDP deflator from FRED (Federal Reserve Economic Data); 2009 is the base year.

Our explanatory variables, x_i, include measures of player production and service time. Total production is measured by Wins Above Replacement (WAR), a well-accepted publicly available measurement of a player's contributions to his team. We include WAR in both their most recent season (pWAR, WAR in the platform year), and total for their career (cWAR, total career WAR excluding the platform year), since previous season and career production are both mentioned as criteria in the arbitration process. A player's service class is included—1 SAE (first time service arbitration eligible), 2 SAE, and 3/4 (third and fourth are typically lumped together) SAE—since the arbitration process is designed to increase player salary each time they go through arbitration.

Attributes that describe a player's skill set and characterize the input production process compose z_{2i}. A player's "Eye" identifies how frequently a player draws a walk.[6] A player's "Bat" identifies how frequently a player gets a hit.[7] A player's "Power" measures their overall power production.[8] These three measures, as implemented in Hakes and Sauer (2007), are simple compared to modern baseball analytic metrics, yet this simplicity likely does a better job capturing how arbitrators characterize player production. Since criteria includes both platform and career performance, we include both aggregate career measures (excluding the platform year) and platform season growth measures.

Other efficiency determinants in z_{2i} include playing time, career consistency, team performance, and player physical defects. A player's plate appearances in the platform year (standardized by taking the player's total and dividing it by the sample maximum) captures the panel's perception of a "showing up to work" effect, in which arbitrators may reward players with higher plate appearances even if player WAR is equal. Career consistency is the percentage of seasons in which a player accumulated within 80% of his best season WAR. Team win percentage in the platform measures overall team performance. Finally, physical defect is captured as the number of days the player spent on the injured list during the platform season.

Table 1 presents summary statistics. Platform measures are indicated with an $\_p$, career measures with an $\_c$, and growth measures with an $\_g$. About half of the sample consists of

[6] $Eye = \frac{UBB+HBP}{PA}$; UBB is unintentional walks; HBP is hit by pitches; PA is plate appearances.

[7] $Bat = \frac{H}{AB}$; H is hits; AB is at-bats (removes walks, HBPs, and sacrifice hits from plate appearances).

[8] $Power = \frac{TB}{H}$; TB is total bases (a single is worth one base, a double two bases, a triple three bases, and a home run four bases). A player with more power accumulates more bases per hit on average.

players that are first-time-arbitration eligible. That drops to 28.5% of the sample for second time arbitration eligible and 22.3% of the sample for third or fourth time arbitration eligible players. If players perform poorly, they are at risk of being released or non-tendered and not advance to the next level of arbitration, explaining why the sample representation drops with increasing service.

Growth measures capture the percentage change in a variable compared to their career numbers. For example, a player with a 1.56 $Power_c$ and a $Power_p$ of 1.70 would have a $Power_g$ of approximately 0.09. Platform measures are likely to be highly correlated with career measures, thus introducing potential collinearity problems when including both career and platform measures. Growth measures, on the other hand, should not be correlated with career measures, while capturing deviations in platform performance.

4.2 Illustration Results

Table 2 presents our base Bayesian and MLE results for the Exponential specifications without bargaining power determinants. Said differently, results show how a player's baseline 'inputs' (platform WAR, career WAR, and service time) impact their 'output' (salary and raise). Unsurprisingly, players with higher pWAR and cWAR earn higher salaries, although platform WAR consistently has a greater impact than career WAR. Players with more service time earn higher salaries, as is designed by the arbitration process. That said, player raises are greatest for 1 SAE players (average \$1.95 million) compared to 2 SAE (\$1.31 million) and 3/4 SAE (\$1.73 million) players, illustrated by negative coefficients for service time variables in the raise specifications.

Our Bayesian approach provides results that are especially similar to those of MLE, with estimates that are within decimal points of each other. Teams seem to generate more surplus through the arbitration process compared to players, with this result more clearly prevalent in the raise specification. If players hate risk more than teams, our results provide potential empirical support for conclusions in the literature that risk aversion leads to lower salary outcomes (Hanany et al. 2007).

Table 3 incorporates efficiency determinants (z_{2i}) to illustrate how player characteristics impact bargaining power in arbitration. As previously illustrated, players with higher pWAR and cWAR earn higher salaries, and salaries increase as service time increases (with reductions in raises after the first time through arbitration). Players relying on their eye attribute for production earn lower salaries in arbitration. This is consistent with MLB's general undervaluing of walks prior to the Moneyball Revolution (see Hakes and Sauer 2006; Hakes and Sauer 2007; Brown et al. 2017). Also consistent with that is the overemphasis of power production, with players relying on power production holding greater bargaining power in arbitration. Career power is positive and statistically significant in the salary and raise specifications. The magnitude of a player's batting average (holding value of production constant) does not seem to impact bargaining power.

Regarding other arbitration case characteristics, players with higher plate appearance totals have greater bargaining power. This may be indicative of an over-reliance on baseball counting stats: players with more plate appearances will score more runs, accumulate more runs batted in, and generate more hits and home runs. Those statistics, which may garner favor with arbitrators, are not necessarily accompanied by an increase in player output.[9]

Player consistency has a negative effect on bargaining power. While this result may seem counterintuitive according to the criteria, there may be a tradeoff between consistent production and high-end production, with arbitrators valuing the potential made evident by a high-performing season. There is also no evidence to suggest that appearing on the injured list decreases bargaining power, although that effect may be captured by the plate appearance variable. Finally, the team win percentage variable is negative in each of the specifications, but is not statistically significant. Again, MLE and Bayesian results are not statistically different.

Table 4 includes growth attributes to account for recent changes to player characteristics. Interestingly, players with higher platform season batting averages have reduced bargaining power. For eye and power, growth coefficients are greater in absolute value in raise specifications than salary specifications, which makes sense given the raise argument emphasizes platform season production. The signs for these different batting attribute coefficients are as expected. Growth and career Eye are negative and statistically significant across specifications, while career power and growth are positive and statistically significant, further supporting earlier results that the arbitration process exhibits familiar pre-Moneyball bi-

[9] The aforementioned counting stats are very teammate reliant. A player will have more runs batted in (RBIs) if there are more runners on base when they come to bat. A player will score more runs if there are quality hitters batting after him.

ases and market inefficiencies. Conclusions are robust to choice of methodology or output specification.[10]

Why might this "biases phenomenon" exist? One explanation relates to the structural nature of baseball arbitration. Baseball arbitration cases rely heavily on analyzing comparable cases—cases that are similar to the player of interest are highlighted during the hearing and negotiation window to help determine salary. While returns to power and on-base ability seemed to correct itself in the free agent market (Hakes and Sauer 2006; Hakes and Sauer 2007), it may be a slower adjustment process for the arbitration market to evolve. Another explanation relates to the baseball knowledge level of the arbitration panel. If arbitrators do not acknowledge the importance of advanced statistics and choose to rely exclusively on traditional stats (which favor power production over on base production), we would expect that to be reflected in negotiated salaries.

5 Conclusions

Both Maximum likelihood estimation and Bayesian estimation are widely used in one-tiered stochastic frontier models. However, for two-tiered stochastic frontier models, maximum likelihood estimation is the predominant estimation method existing in the literature. We fill this clear gap in the literature by proposing an alternative estimation method—Bayesian approach—for estimating 2TSF models with standard and mostly widely used specifications—Exponential specifications—and also derive the corresponding results for 2TSF with Half-Normal specifications. Moreover, this approach has the added flexibility of incorporating efficiency determinants, which can be useful in empirical applications like the bargaining situation illustrated in this paper. Further, the posterior distributions for parameters in two-tiered stochastic frontier model are derived, and thus the Bayesian approach can be easily deployed using Gibbs sampling method or Metropolis' algorithm.

We illustrate the applicability of our Bayesian approaches for the 2TSF using baseball arbitration data and also compare the results with those from maximum likelihood estimation. Consistent with pre-Moneyball market characteristics, we find that players who provide value with power production have greater bargaining position and earn higher salaries in baseball

[10] Results are also robust to the distributional assumption. Bayesian results using the Half-Normal in lieu of the Exponential are available upon request.

arbitration, while players who generate value with on-base abilities have less bargaining power and earn lower salaries in baseball arbitration. These results are robust to the output measures used (salary or raise). It is also noteworthy that our Bayesian approach provides similar results to MLE in our empirical application. There are pros and cons in choosing Bayesian techniques over MLE, but our empirical results suggest that the conclusions drawn are generally robust across techniques. However these similar findings may not necessarily hold in other empirical applications.

Future work could focus on Bayesian estimation of the setting with other variants of 2TSF, such as Truncated-Normal or Generalized Exponential specifications. Future work could also use the posterior odds ratio to determine which parametric specification (Exponential, Half-Normal, Generalized Exponential, etc) is suitable. Another interesting strand of future work is to compare the 2TSF with a deterministic double frontier approach (Hadley and Ruggiero 2006), which is constructed using non-parametric methods—Data Envelopment Analysis (DEA) or Free Disposal Hull (FDH). This comparison would be in parallel to some literature that has compared one-tiered stochastic frontier with DEA or FDH, such as Gong and Sickles (1992). Another interesting path would be to consider how to optimally combine 2TSF with non-parametric double frontier approach, similarly as those discussions combining one-tiered stochastic frontier model with DEA (Parmeter and Zelenyuk 2019; Tsionas 2021).

Appendix

In this appendix, similar to the case of the Exponential specification, we derive the corresponding Bayesian results for the 2TSF models with the Half-Normal specification.

The Half-Normal Specification Without Efficiency Determinants

The Half-Normal 2TSF has the Normal-'Half-Normal'-'Half-Normal' distribution, i.e., $v_i \sim N(0, 1/\phi)$, $u_i \sim N_+(0, 1/\phi_1)$, and $w_i \sim N_+(0, 1/\phi_2)$. The conditional density for y_i has a closed form, see Papadopoulos (2015).

We specify the following prior for the parameters,

$$\beta \propto \text{cons}, \phi \sim \mathcal{G}(a_0, b_0), \phi_1 \sim \mathcal{G}(a_1, b_1), \phi_2 \sim \mathcal{G}(a_2, b_2). \tag{A.1}$$

The conditional posterior distributions for β and ϕ are the same as (3.3) and (3.4). Conditional posterior distributions for the remaining parameters are,

$$\phi_1 | u \sim \mathcal{G}\left(a_1 + \frac{n}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{n} u_i^2\right), \tag{A.2}$$

$$\phi_2 | w \sim \mathcal{G}\left(a_2 + \frac{n}{2}, b_2 + \frac{1}{2}\sum_{i=1}^{n} w_i^2\right), \tag{A.3}$$

$$u_i | y, x, w, \beta, \phi, \phi_1 \sim N\left(\frac{\phi}{\phi + \phi_1}(x_i\beta + w_i - y_i), \frac{1}{\phi + \phi_1}\right) I(u_i \geq 0), \tag{A.4}$$

and

$$w_i | y, x, u, \beta, \phi, \phi_2 \sim N\left(\frac{\phi}{\phi + \phi_2}(y_i + u_i - x_i\beta), \frac{1}{\phi + \phi_2}\right) I(w_i \geq 0). \tag{A.5}$$

The Half-Normal Specification With Efficiency Determinants

In the Half-Normal 2TSF with efficiency determinants, we specify

$$\begin{aligned} y_i =& x_i\beta + v_i - u_i + w_i, \\ & u_i \sim N_+(0, \sigma_{ui}^2), \sigma_{ui}^2 = \exp(z_{1i}\delta_1), \\ & w_i \sim N_+(0, \sigma_{wi}^2), \sigma_{wi}^2 = \exp(z_{2i}\delta_2), \\ & v_i \sim N(0, 1/\phi). \end{aligned} \tag{A.6}$$

The effects of z_{1i} on u_i are achieved through affecting its mean value $\sqrt{2/\pi}\sigma_{ui}$, while the effects of z_{2i} on w_i are achieved through $\sqrt{2/\pi}\sigma_{wi}$. We specify the following prior for the parameters,

$$\beta \propto \text{cons}, \phi \sim \mathcal{G}(a_0, b_0), \delta_1 \sim N_q(0, R_1), \delta_2 \sim N_r(0, R_2). \tag{A.7}$$

The conditional posterior distributions for β and ϕ are equivalent to (3.3) and (3.4). The conditional posterior distributions for the remaining parameters are,

$$\delta_1|z_1, u \sim \prod_{i=1}^{n} \Big(\exp(z_{1i}\delta_1)^{-\frac{1}{2}} \exp(-\frac{1}{2}\exp(z_{1i}\delta_1)^{-1}u_i^2)\Big) \exp(-\frac{\delta_1' R_1^{-1}\delta_1}{2}), \tag{A.8}$$

$$\delta_2|z_2, w \sim \prod_{i=1}^{n} \Big(\exp(z_{2i}\delta_2)^{-\frac{1}{2}} \exp(-\frac{1}{2}\exp(z_{2i}\delta_2)^{-1}w_i^2)\Big) \exp(-\frac{\delta_2' R_2^{-1}\delta_2}{2}), \tag{A.9}$$

$$u_i|y, x, z_1, w, \beta, \phi, \delta_1 \sim N\left(\frac{\phi}{\phi + \exp(z_{1i}\delta_1)^{-1}}(x_i\beta + w_i - y_i), \frac{1}{\phi + \exp(z_{1i}\delta_1)^{-1}}\right) I(u_i \geq 0), \tag{A.10}$$

and

$$w_i|y, x, z_2, u, \beta, \phi, \delta_2 \sim N\left(\frac{\phi}{\phi + \exp(z_{2i}\delta_2)^{-1}}(y_i + u_i - x_i\beta), \frac{1}{\phi + \exp(z_{2i}\delta_2)^{-1}}\right) I(w_i \geq 0). \tag{A.11}$$

When efficiency determinants exist for w_i, but not u_i, we assume $u_i \sim N_+(0, 1/\phi_1)$. Further, assume the prior for $\phi_1 \sim \mathcal{G}(a_1, b_1)$. The conditional posterior distributions for β, ϕ, δ_2 and w_i are unchanged and are listed in (3.3), (3.4), (A.9), and (A.11). The conditional posterior distributions for ϕ_1 and u_i are

$$\phi_1|u \sim \mathcal{G}\left(a_1 + \frac{n}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{n} u_i^2\right), \tag{A.12}$$

and

$$u_i|y, x, w, \beta, \phi, \phi_1 \sim N\left(\frac{\phi}{\phi + \phi_1}(x_i\beta + w_i - y_i), \frac{1}{\phi + \phi_1}\right) I(u_i \geq 0). \tag{A.13}$$

On the other hand, it is possible that there are only efficiency determinants for u_i, but not for w_i. For this situation, we assume $w_i \sim N_+(0, 1/\phi_2)$. Further, assume the prior for $\phi_2 \sim \mathcal{G}(a_2, b_2)$. The conditional posterior distributions for β, ϕ, δ_1 and u_i are unchanged and are listed in (3.3), (3.4), (A.8), and (A.10). Conditional posterior distributions for ϕ_2 and w_i are

$$\phi_2|w \sim \mathcal{G}\left(a_2 + \frac{n}{2}, b_2 + \frac{1}{2}\sum_{i=1}^{n} w_i^2\right), \tag{A.14}$$

and

$$w_i|y, x, u, \beta, \phi, \phi_2 \sim N\left(\frac{\phi}{\phi+\phi_2}(y_i+u_i-x_i\beta), \frac{1}{\phi+\phi_2}\right) I(w_i \geq 0), \tag{A.15}$$

respectively.

References

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21–37.

Blanco, G. (2017), Who benefits from job placement services? a two-sided analysis, *Journal of Productivity Analysis* 47, 33–47.

Brown, D. T., C. R. Link, and S. L. Rubin (2017), Moneyball after 10 years: how have major league baseball salaries adjusted?, *Journal of Sports Economics* 18, 771–786.

Chawla, M. (2002), Estimating the extent of patient ignorance, in *World Bank Economists' Forum*, volume 2, World Bank Publications, p. 3.

Das, T. and S. W. Polachek (2017), Estimating labor force joiners and leavers using a heterogeneity augmented two-tier stochastic frontier, *Journal of Econometrics* 199, 156–172.

Dworkin, J. B. (1977), Final position arbitration and intertemporal compromise, *Relations industrielles/Industrial Relations* 32, 250–261.

Farmer, A., P. Pecorino, and V. Stango (2004), The causes of bargaining failure: Evidence from major league baseball, *The Journal of Law and Economics* 47, 543–568.

Faurot, D. J. (2001), Equilibrium explanation of bargaining and arbitration in major league baseball, *Journal of Sports Economics* 2, 22–34.

Faurot, D. J. and S. McAllister (1992), Salary arbitration and pre-arbitration negotiation in major league baseball, *ILR Review* 45, 697–710.

Ferona, A. and E. G. Tsionas (2012), Measurement of excess bidding in auctions, *Economics Letters* 116, 377–380.

Fizel, J. (1996), Bias in salary arbitration: the case of major league baseball, *Applied Economics* 28, 255–265.

Fizel, J., A. C. Krautmann, and L. Hadley (2002), Equity and arbitration in major league baseball, *Managerial and Decision Economics* 23, 427–435.

Fried, H. O. and L. W. Tauer (2019), Efficient wine pricing using stochastic frontier models, *Journal of Wine Economics* 14, 164–181.

Gaynor, M. and S. W. Polachek (1994), Measuring information in the market: An application to physician services, *Southern Economic Journal* , 815–831.

Geman, S. and D. Geman (1984), Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Transactions on pattern analysis and machine intelligence* , 721–741.

Gong, B.-H. and R. C. Sickles (1992), Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal of econometrics* 51, 259–284.

Greene, W. (2005), Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of econometrics* 126, 269–303.

Griffin, J. E. and M. F. Steel (2004), Semiparametric bayesian inference for stochastic frontier models, *Journal of econometrics* 123, 121–152.

— (2007), Bayesian stochastic frontier analysis using winbugs, *Journal of Productivity Analysis* 27, 163–176.

Groot, W. and H. Oosterbeek (1994), Stochastic reservation and offer wages, *Labour Economics* 1, 383–390.

Gustafson, E. and L. Hadley (1995), Arbitration and salary gaps in major league baseball, *Quarterly Journal of Business and Economics* , 32–46.

Hadley, L. and J. Ruggiero (2006), Final-offer arbitration in major league baseball: A nonparametric analysis, *Annals of Operations Research* 145, 201–209.

Hakes, J. K. and R. D. Sauer (2006), An economic evaluation of the moneyball hypothesis, *Journal of Economic Perspectives* 20, 173–186.

— (2007), The moneyball anomaly and payroll efficiency: A further investigation., *International Journal of Sport Finance* 2.

Hanany, E., D. M. Kilgour, and Y. Gerchak (2007), Final-offer arbitration and risk aversion in bargaining, *Management Science* 53, 1785–1792.

Holmes, P. M., R. Simmons, and D. J. Berri (2018), Moneyball and the baseball players' labor market, *International Journal of Sport Finance* 13, 141–155.

Hu, Z. and K. Pei (2020), Bi-directional R&D spillovers and operating performance: A two-tier stochastic frontier model, *Economics Letters* 195, 109485.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982), On the estimation of technical inefficiency in the stochastic frontier production model, *Journal of Econometrics* 19, 233–238.

Jradi, S., C. F. Parmeter, and J. Ruggiero (2019), Quantile estimation of the stochastic frontier model, *Economics Letters* 182, 15–18.

— (2021), Quantile estimation of stochastic frontiers with the normal-exponential specification, *European Journal of Operational Research* 295, 475–483.

Kahn, L. M. (2000), The sports business as a labor market laboratory, *Journal of economic perspectives* 14, 75–94.

Kim, Y. and P. Schmidt (2000), A review and empirical comparison of bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data, *Journal of productivity Analysis* 14, 91–118.

Klein, N., H. Herwartz, and T. Kneib (2020), Modelling regional patterns of inefficiency: A bayesian approach to geoadditive panel stochastic frontier analysis with an application to cereal production in england and wales, *Journal of Econometrics* 214, 513–539.

Koop, G., M. F. Steel, et al. (2001), Bayesian analysis of stochastic frontier models, *A companion to theoretical econometrics* 1, 520–73.

Kumbhakar, S. C., S. Ghosh, and J. T. McGuckin (1991), A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms, *Journal of Business & Economic Statistics* 9, 279–286.

Kumbhakar, S. C. and C. F. Parmeter (2009), The effects of match uncertainty and bargaining on labor market outcomes: evidence from firm and worker specific estimates, *Journal of Productivity Analysis* 31, 1–14.

— (2010), Estimation of hedonic price functions with incomplete information, *Empirical Economics* 39, 1–25.

Kumbhakar, S. C., C. F. Parmeter, and V. Zelenyuk (2020a), Stochastic Frontier Analysis: Foundations and Advances I, in S. C. Ray, R. Chambers, and S. Kumbhakar, eds., *Handbook of Production Economics*, Singapore: Springer, pp. 1–39.

— (2020b), Stochastic Frontier Analysis: Foundations and Advances II, in S. C. Ray, R. Chambers, and S. Kumbhakar, eds., *Handbook of Production Economics*, Singapore: Springer, pp. 1–38.

Kumbhakar, S. C. and M. G. Tsionas (2021), Dissections of input and output efficiency: A generalized stochastic frontier model, *International Journal of Production Economics* 232, 107940.

Liu, Y., X. Yao, and T. Wei (2019), Energy efficiency gap and target setting: A study of information asymmetry between governments and industries in china, *China Economic Review* 57, 101341.

Meeusen, W. and J. van den Broeck (1977), Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435–444.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *The journal of chemical physics* 21, 1087–1092.

Miller, P. A. (2000), A theoretical and empirical comparison of free agent and arbitration-eligible salaries negotiated in major league baseball, *Southern Economic Journal* 67, 87–104.

Murphy, A. and E. Strobl (2008), Employer and employee ignorance in developing countries: The case of trinidad and tobago, *Review of Development Economics* 12, 339–353.

Papadopoulos, A. (2015), The half-normal specification for the two-tier stochastic frontier model, *Journal of Productivity Analysis* 43, 225–230.

— (2018), The two-tier stochastic frontier framework (2TSF): Measuring frontiers wherever they may exist, in *North American Productivity Workshop*, Springer, pp. 163–194.

— (2021a), Measuring the effect of management on production: a two-tier stochastic frontier approach, *Empirical Economics* 60, 3011–3041.

— (2021b), Stochastic frontier models using the generalized exponential distribution, *Journal of Productivity Analysis* 55, 15–29.

— (2022), The Nash bargaining two-tier stochastic frontier model. Unpublished working paper.

Papadopoulos, A. and C. F. Parmeter (2022), Quantile methods for stochastic frontier analysis. Unpublished working paper.

Papadopoulos, A., C. F. Parmeter, and S. C. Kumbhakar (2021), Modeling dependence in two-tier stochastic frontier models, *Journal of Productivity Analysis* 56, 85–101.

Parmeter, C. F. (2018), Estimation of the two-tiered stochastic frontier model with the scaling property, *Journal of Productivity Analysis* 49, 37–47.

Parmeter, C. F. and V. Zelenyuk (2019), Combining the virtues of stochastic frontier and data envelopment analysis, *Operations Research* 67, 1628–1658.

Pinheiro, R. and S. Szymanski (2022), All runs are created equal: Labor market efficiency in major league baseball, *Journal of Sports Economics* , 15270025221085712.

Polachek, S. W. and B. J. Yoon (1987), A two-tiered earnings frontier estimation of employer and employee information in the labor market, *The Review of Economics and Statistics* 69, 296.

— (1996), Panel estimates of a two-tiered earnings frontier, *Journal of Applied Econometrics* 11.

Pu, G., Y. Zhang, and L.-C. Chou (2022), Estimating financial information asymmetry in real estate transactions in china-an application of two-tier frontier model, *Information Processing & Management* 59, 102860.

Reifschneider, D. and R. Stevenson (1991), Systematic departures from the frontier: A framework for the analysis of firm inefficiency, *International Economic Review* 32, 715–723.

Scully, G. W. (1974), Pay and performance in major league baseball, *The American Economic Review* 64, 915–930.

Sharif, N. R. and A. A. Dar (2007), An empirical investigation of the impact of imperfect information on wages in canada, *Review of Applied Economics* 3, 137–155.

Tomini, S., W. Groot, and M. Pavlova (2012), Paying informally in the albanian health care sector: a two-tiered stochastic frontier model, *The European Journal of Health Economics* 13, 777–788.

Tsionas, M. G. (2020), Quantile Stochastic Frontiers, *European Journal of Operational Research* 282, 1177–1184.

— (2021), Optimal combinations of stochastic frontier and data envelopment analysis models, *European Journal of Operational Research* 294, 790–800.

Tsionas, M. G., A. G. Assaf, and A. Andrikopoulos (2020), Quantile stochastic frontier models with endogeneity, *Economics Letters* 188, 108964.

Tsionas, M. G. and S. C. Kumbhakar (2021), Stochastic frontier models with time-varying conditional variances, *European Journal of Operational Research* 292, 1115–1132.

Tsionas, M. G. and S. K. Mallick (2019), A bayesian semiparametric approach to stochastic frontiers and productivity, *European Journal of Operational Research* 274, 391–402.

Van den Broeck, J., G. Koop, J. Osiewalski, and M. F. Steel (1994), Stochastic frontier models: A bayesian perspective, *Journal of econometrics* 61, 273–303.

Verteramo Chiu, L. J., L. W. Tauer, and Y. T. Gröhn (2022), Pricing efficiency in livestock auction markets: A two-tier frontier approach, *Agricultural Economics* .

Wang, P.-Y. (2017), Six component panel stochastic model. Master Thesis. National Taiwan University.

Wang, Y. (2018), Bargaining matters: An analysis of bilateral aid to developing countries, *Journal of International Relations and Development* 21, 1–21.

Wittman, D. (1986), Final-offer arbitration, *Management Science* 32, 1551–1561.

Xu, C., L. Yang, B. Zhang, and M. Song (2021), Bargaining power and information asymmetry in China's water market: an empirical two-tier stochastic frontier analysis, *Empirical Economics* 61, 2395–2418.

Zhang, H., J. Zhang, Y. Yang, and Q. Zhou (2018), Bargaining power in tourist shopping, *Journal of Travel Research* 57, 947–961.

Zhao, S. (2021), Quantile estimation of stochastic frontier models with the normal–half normal specification: A cumulative distribution function approach, *Economics Letters* 206, 109998.

Table 1: Summary Statistics

Variable	Min	Q1	Median	Mean	Q3	Max
Salary ($)	358196.12	1349345.34	2568738.45	3034956.05	3930961.01	15003823.56
Log of Salary	12.7890	14.1150	14.7590	14.6870	15.1840	16.5240
Raise ($)	11530.17	924651.20	1568039.27	1722383.29	2333680.39	7604747.13
Log of Raise	9.3530	13.7370	14.2650	14.0980	14.6630	15.8440
pWAR	-1.6000	0.6000	1.6000	1.8400	2.7000	9.4000
cWAR	-5.3000	1.3000	3.1000	4.0200	5.8500	31.0000
I(serviceTime=2)	0.0000	0.0000	0.0000	0.2850	1.0000	1.0000
I(serviceTime=3/4)	0.0000	0.0000	0.0000	0.2230	0.0000	1.0000
Eye_g	-1.0000	-0.1400	0.0180	0.0520	0.1980	3.0520
Eye_c	0.0270	0.0670	0.0830	0.0850	0.1010	0.1680
Bat_g	-1.0000	-0.0620	0.0170	0.0190	0.0910	1.7030
Bat_c	0.1780	0.2480	0.2620	0.2610	0.2760	0.3240
Power_g	-1.0000	-0.0490	0.0160	0.0190	0.0820	0.7930
Power_c	1.1580	1.4450	1.5670	1.5660	1.6760	2.0350
PA_p/max(PA_p)	0.0000	0.4400	0.6260	0.6040	0.7950	1.0000
Consistency_c	0.0000	0.2000	0.2500	0.2950	0.4000	1.0000
DLdays_p	0.0000	0.0000	0.0000	29.7540	53.5000	192.0000
tmWinPct	0.3150	0.4570	0.5120	0.5080	0.5560	0.7160

Table 2: Log Salary/Raise Regression Without Inefficiency Determinants

	Bayesian	MLE	Bayesian	MLE
Outputs (y_i)	Salary	Salary	Raise	Raise
Constant	13.954***	13.980***	14.221***	14.250***
	(0.063)	(0.072)	(0.06)	(0.079)
pWAR	0.143***	0.140***	0.191***	0.188***
	(0.011)	(0.011)	(0.014)	(0.014)
cWAR	0.069***	0.068***	0.029***	0.029***
	(0.005)	(0.005)	(0.006)	(0.006)
I(serviceTime==2)	0.353***	0.364***	−0.472***	−0.446***
	(0.041)	(0.040)	(0.052)	(0.051)
I(serviceTime==3/4)	0.563***	0.590***	−0.338***	−0.336***
	(0.048)	(0.049)	(0.057)	(0.057)
ϕ	7.24***	8.615***	9.262***	9.264***
	(0.859)	(1.857)	(1.533)	(1.823)
λ_1	0.195***	0.231***	0.531***	0.534***
	(0.036)	(0.044)	(0.037)	(0.037)
λ_2	0.162***	0.174***	0.151***	0.116**
	(0.032)	(0.048)	(0.029)	(0.056)
		—Surplus Extracted—		
Players: $E(w)$	0.160	0.173	0.15	0.117
Teams: $E(u)$	0.193	0.23	0.53	0.532
Net Surplus: $E(w-u)$	−0.033	−0.058	−0.38	−0.416
Players: $1-E(e^{-w})$	0.138	0.147	0.13	0.104
Teams: $1-E(e^{-u})$	0.162	0.188	0.343	0.344
Net Surplus: $E(e^{-u}-e^{-w})$	−0.024	−0.04	−0.213	−0.239

Note: Statistical significance denoted $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3: Log Salary/Raise Regression Without Attribute Growth Variables

	Bayesian	MLE	Bayesian	MLE
Outputs (y_i)	Salary	Salary	Raise	Raise
Constant	13.939***	14.228***	14.139***	14.133***
	(0.047)	(0.061)	(0.053)	(0.056)
pWAR	0.126***	0.124***	0.180***	0.180***
	(0.011)	(0.013)	(0.014)	(0.014)
cWAR	0.072***	0.064***	0.034***	0.033***
	(0.005)	(0.006)	(0.006)	(0.007)
I(serviceTime==2)	0.340***	0.194***	−0.539***	−0.528***
	(0.037)	(0.046)	(0.052)	(0.052)
I(serviceTime==3/4)	0.523***	0.225***	−0.430***	−0.444***
	(0.045)	(0.058)	(0.059)	(0.061)
—Inefficiency Determinants (z_{2i})—				
Constant	−7.506***	−5.342***	−5.413***	−5.749***
	(0.793)	(1.478)	(0.928)	(1.643)
Eye_c	−15.929***	−11.329***	−11.334***	−11.367***
	(3.084)	(3.411)	(3.368)	(3.563)
Bat_c	1.988	−1.043	−1.061	−1.034
	(2.656)	(4.375)	(2.723)	(4.812)
Power_c	2.738***	2.012***	1.900***	2.156***
	(0.288)	(0.485)	(0.387)	(0.538)
PA_p/max(PA_p)	4.508***	4.876***	4.837***	4.628***
	(0.549)	(0.615)	(0.638)	(0.668)
Consistency_c	−1.308**	−1.708***	−1.716***	−1.522**
	(0.569)	(0.580)	(0.675)	(0.641)
DLdays_p	0.003	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
tmWinPct	−1.006	−1.715	−1.756	−1.598
	(1.087)	(1.183)	(1.181)	(1.270)
ϕ	11.132***	12.394***	12.405***	12.895***
	(1.355)	(1.406)	(1.958)	(1.964)
λ_1	0.187***	0.411***	0.463***	0.461***
	(0.029)	(0.046)	(0.030)	(0.031)
—Surplus Extracted—				
Players: $E(w)$	0.201	0.237	0.203	0.212
Teams: $E(u)$	0.185	0.359	0.461	0.461
Net Surplus: $E(w-u)$	0.016	−0.122	−0.258	−0.249
Players: $1-E(e^{-w})$	0.157	0.18	0.157	0.163
Teams: $1-E(e^{-u})$	0.156	0.273	0.311	0.311
Net Surplus: $E(e^{-u}-e^{-w})$	0.001	−0.092	−0.154	−0.148

Note: Statistical significance denoted $^{***}\ p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4: Log Salary/Raise Regression With Attribute Growth Variables

	Bayesian	MLE	Bayesian	MLE
Outputs (y_i)	Salary	Salary	Raise	Raise
		— Inputs (x_i) —		
Constant	13.900***	14.254***	14.137***	14.118***
	(0.047)	(0.065)	(0.054)	(0.056)
pWAR	0.141***	0.125***	0.194***	0.195***
	(0.011)	(0.014)	(0.016)	(0.016)
cWAR	0.070***	0.061***	0.026***	0.026***
	(0.005)	(0.006)	(0.007)	(0.007)
I(serviceTime=2)	0.341***	0.226***	−0.520***	−0.521***
	(0.038)	(0.044)	(0.051)	(0.051)
I(serviceTime=3/4)	0.537***	0.302***	−0.391***	−0.399***
	(0.047)	(0.054)	(0.057)	(0.060)
		— Efficiency Determinants (z_{2i}) —		
Constant	−5.212***	−8.800***	−8.869***	−8.646***
	(1.029)	(1.987)	(0.925)	(1.992)
Eye_g	−0.871***	−1.102***	−1.120***	−1.249***
	(0.284)	(0.354)	(0.366)	(0.347)
Eye_c	−19.068***	−16.238***	−16.242***	−16.285***
	(3.435)	(4.188)	(3.226)	(3.923)
Bat_g	−2.021***	−1.928**	−1.938*	−1.923**
	(0.705)	(0.954)	(1.021)	(0.985)
Bat_c	−2.306	3.787	3.771	3.793
	(1.927)	(5.389)	(2.409)	(5.283)
Power_g	0.883	1.942*	1.941**	1.972**
	(0.816)	(1.010)	(0.900)	(0.927)
Power_c	2.481***	3.175***	3.071***	3.076***
	(0.515)	(0.611)	(0.485)	(0.619)
PA_p/max(PA_p)	4.323***	5.080***	5.043***	4.682***
	(0.502)	(0.710)	(0.871)	(0.668)
Consistency_c	−1.253**	−1.481**	−1.486**	−1.341**
	(0.577)	(0.667)	(0.683)	(0.640)
DLdays_p	0.003*	−0.006	0.001	0.001
	(0.002)	(0.004)	(0.002)	(0.002)
tmWinPct	−1.573	−0.910	−0.947	−0.741
	(1.010)	(1.356)	(1.125)	(1.295)
ϕ	11.325***	12.959***	12.971***	13.465***
	(1.398)	(1.493)	(1.912)	(1.985)
λ_1	0.180***	0.421***	0.461***	0.458***
	(0.030)	(0.048)	(0.030)	(0.031)
		— Surplus Extracted —		
Players: $E(w)$	0.209	0.197	0.195	0.209
Teams: $E(u)$	0.178	0.366	0.460	0.456
Net Surplus: $E(w-u)$	0.030	−0.169	−0.266	−0.247
Players: $1-E(e^{-w})$	0.162	0.151	0.149	0.16
Teams: $1-E(e^{-u})$	0.151	0.276	0.311	0.309
Net Surplus: $E(e^{-u}-e^{-w})$	0.011	−0.126	−0.161	−0.149

Note: Statistical significance denoted $^{***}\ p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$