May 2018

Production, efficiency and managerial selection of workers into peer networks

Hyunseok Jung
Syracuse University

Follow this and additional works at: https://surface.syr.edu/etd

Part of the Social and Behavioral Sciences Commons

Recommended Citation
https://surface.syr.edu/etd/856
ABSTRACT

This dissertation develops empirical models that account for worker interactions, managerial selectivity, and technical inefficiency in the production process.

The first chapter, entitled “Stochastic Frontier Models with Network Selectivity,” develops a model where workers produce output through peer-effect networks, while managerial selectivity of workers affects worker inefficiency. The intuition behind this model is that managers may consider optimal combinations of workers to produce the best results, and this selectivity in the worker network may affect worker productivity.

The second chapter, entitled “Network Competition and Team Chemistry in the NBA,” models simultaneous interactions between multiple networks where agents cooperate with peers within their own networks but compete with non-peers from other networks. This paper presents the first econometric model to consider multiple peer networks where workers are engaged in simultaneous competition around a single outcome variable.

Lastly, the third chapter, entitled “Adaptive LASSO for Stochastic Frontier Models with Many Efficient Firms,” develops a procedure to select a subset of maximally efficient firms in the sample of interest. In this model, firm inefficiency is measured as a distance from an estimated optimal production level, and I apply the LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996) to identify a subset of firms whose inefficiencies are estimated as exactly zero. This methodology can be applied to any classification problem where our interest is to identify a subset of best (worst) individuals among a large number of candidates.
Production, efficiency and managerial selection of workers into peer networks

by

Hyunseok Jung

B.A., Seoul National University, S. Korea, 2008
M.A., Korea Development Institute, S. Korea, 2013
MPA, Syracuse University, 2014

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics

Syracuse University

May 2018
For my wife, Jaewon.
Acknowledgments

This dissertation would never have been possible without the support of my enthusiastic advisor, Dr. William Horrace. He has made every possible effort for my success in academia. He always tried to see my potential beyond what I had and showed me what it means to be a dedicated.

I am also indebted to the other members of my dissertation committee, Dr. Alfonso Flores-Lagunes and Yoonseok Lee. I am hugely appreciative to their time spent reading my papers and their advises for improving them.

Perhaps the greatest debt is that I owe to my wife, Jeawon. I will always remember that your sacrifice and support made this journey possible. You are the most important person in my world and I dedicate this thesis to you.

Syracuse, New York
April, 2018
Contents

List of Figures ix

List of Tables x

1 Stochastic Frontier Models with Network Selectivity 1
  1.1 Introduction .................................................. 2
  1.2 Model and Estimation ........................................ 4
    1.2.1 Econometric model .................................. 4
    1.2.2 Estimation of the Stochastic Frontier ............... 8
    1.2.3 A Two-Step Approach ................................ 9
    1.2.4 Estimation of Individual Inefficiencies ............ 10
  1.3 Empirical Application ...................................... 11
    1.3.1 Data and Variables .................................. 11
    1.3.2 Estimation Results .................................... 15
  1.4 Conclusions ................................................ 19

References 21

2 Network Competition and Team Chemistry in the NBA 25
  2.1 Introduction .................................................. 26
  2.2 Model and estimation ....................................... 29
    2.2.1 Econometric model .................................. 29
    2.2.2 Estimation ............................................. 35

vi
2.3 Empirical application .............................................. 39
  2.3.1 Empirical model and variables .............................. 39
  2.3.2 Estimation strategy ........................................... 43
  2.3.3 (Preliminary) Estimation results .......................... 44

2.4 Conclusion .......................................................... 48

References ............................................................... 49

3 Adaptive LASSO for SFM with Many Efficient Firms 54

  3.1 Introduction ....................................................... 55
  3.2 Adaptive LASSO for many efficient firms .................... 59
    3.2.1 Production Function ....................................... 59
    3.2.2 Adaptive LASSO estimator ................................. 61
  3.3 Computation ...................................................... 65
    3.3.1 Optimization algorithm ................................... 65
    3.3.2 Tuning parameter .......................................... 67
  3.4 Asymptotic Theory ............................................... 68
  3.5 Simulations ....................................................... 72
    3.5.1 Setup ........................................................ 72
    3.5.2 Results ...................................................... 73
  3.6 Empirical application ........................................... 77
    3.6.1 Comparison with "Ranking and Selection" (R&S) procedure ........................................... 77
    3.6.2 Results ...................................................... 78
  3.7 Conclusion ........................................................ 79

References ............................................................... 83

vii
Appendices  88

A Stochastic Frontier Models with Network Selectivity  89
  A.1 Mathematical Derivations .............................................. 89
  A.2 Three-step Estimation Approach .................................... 91
  A.3 The partial and the second derivatives of the log-likelihood and adjustment of
      the standard error ...................................................... 93

B Network Competition and Team Chemistry in the NBA  99
  B.1 GMM for the heterogeneous network interaction model .......... 99
  B.2 Identification condition for the equation (2.11) .................. 102
  B.3 Lemma B.3.1 .................................................................. 103
  B.4 Derivation of the best IV and the $F_i$s, and its asymptotic distribution .... 105
  B.5 Descriptive Statistics for NBA Data .................................. 108
  B.6 Estimation results .......................................................... 112

C Adaptive LASSO for SFM with Many Efficient Firms  116
  C.1 Proofs of the lemmas and the theorem ............................... 116
  C.2 The equivalence between the one step and the two step estimation .... 124
  C.3 The proximity of the new and the standard algorithm ............ 126
  C.4 New algorithm ............................................................ 127
  C.5 Simulation results for $\beta$ .............................................. 129

Vita  130
List of Figures

1.1 (a) Bivariate exponential distribution of $u$ and $\epsilon$; (b) $E(u_{ist}|\epsilon_{st} < \delta_{st})$ by the value of $\alpha$ when $\sigma_u = 1$ ......................... 8

2.1 Wins Produced per Minute per Player for Utah Jazz, 2013-14 ................. 46

3.1 PDFs of inefficiency with different $\sigma_u$ values and an example of draws from each PDF. ................................................................. 73

3.2 Distribution of estimates for $\alpha_0$ from LASSO and LSDV ...................... 75

3.3 Visualization of the selection performance ($N = 100$) .......................... 76

3.4 Distribution of the rice farm inefficiencies ............................................. 78

3.5 Firm IDs estimated as fully efficient by LASSO or included in the best subset by R&S ................................................................. 79

C.1 Model (1) .................................................................................. 126

C.2 Model (2) .................................................................................. 127
List of Tables

1.1 Descriptive statistics .................................................. 14
1.2 Estimation of the selection equation ................................. 16
1.3 Three step ML estimation of the outcome function .............. 17
2.1 Ranked Peer-Effects Based on Wins Produced by Season ...... 45
3.1 Estimation accuracy for \( \hat{U}(\Pi) \) ................................ 81
3.2 Selection accuracy for \( \hat{U}(\Pi) \) .................................. 82
B.1 Season 2013-14 ....................................................... 109
B.2 Season 2014-15 ....................................................... 110
B.3 Season 2015-16 ....................................................... 111
B.4 2013-14 Season Estimates ........................................... 113
B.5 2014-15 Season Estimates ........................................... 114
B.6 2014-15 Season Estimates ........................................... 115
C.1 Simulation results for \( \beta \) ........................................ 129
We allow for worker-level inefficiency that is correlated with the manager’s selection equation in the network production function model of Horrace, Liu and Patacchini (2016), which transforms labor into productive output through peer-effect networks selected by the manager. The proposed specification is a "composed error" version of their model where managerial selectivity works through the worker-level inefficiency term instead of the noise term. The new specification captures both worker-level inefficiency and the manager’s ability to efficiently select teams to produce output. As the correlation between the manager’s selection equation and worker inefficiency goes to zero, our parametric model reduces to the normal-exponential stochastic frontier model of Aigner, Lovell and Schmidt (1977) with peer-effects. As the variance of worker inefficiency goes to zero, the manager’s choices become random. A brief application to the NBA is provided.
1.1 Introduction

Horrace, Liu and Patacchini (2016) develop a network production model where a manager selects workers into teams (peer-effect networks) and where the productivity of each worker affects the productivities of teammates and vice versa. This induces a production function selection bias that can be corrected by augmenting the production equation with a poly-chotomous team selection equation. Estimation of the model is at the worker-level within a single firm and proceeds as a two-stage process. The first stage models the manager’s optimal team choice, based on the aggregate performance of each possible team that she can create from the pool of workers. In this model a potential team with better aggregate performance (output) has a higher probability of being selected by the manager in the first-stage. This yields a selection bias term (a positive inverse Mill’s ratio) that is used as a correction in estimation of the second-stage production function. Depending on the sign of the correlation between the selection equation error and the production function error, optimal managerial choice may either increase or decrease output. Horrace et. al. (2016) claim that this bias term can be interpreted as managerial inefficiency.

However, the source of variability in the production function is at the worker-level, so we would like to consider models that also account for worker-level inefficiency (as opposed to only managerial inefficiency), and propose a network production specification with a composed error: a two-sided component that represents statistical noise and a one-sided component that represents worker inefficiency and can only subtract from the worker’s (and the team’s) productivity. In our model the manager may have information on the inefficiency draws of individual workers and makes her choices based on this knowledge, so any bias induced in the model is strictly though the worker inefficiency component. In a world where production is based on a composed error it is not unreasonable to suspect that the manager only has information on one of the components.
Another distinctive feature of our model is that the effect of the manager’s selection on the productivities of the workers is in one direction: positive. In Horrace et al. (2016), depending on the sign of the correlation between the selection errors and the production errors, the manager’s choices can have either positive or negative effects on worker performance, and their model accommodates the (perhaps) unrealistic situation where the expected performance of an optimally selected team of workers is lower than a feasible alternative. That is, the probability of a lineup being selected may be high, even though its expected performance is low. The proposed specification ensures that manager’s choices are consistent with the goal of maximizing output at the cost of restricting the magnitude of the correlation between the selection equation and the production function, as we shall see.

Kumbhakar et al. (2009) and Greene (2010) deal with the sample selection issue in the stochastic frontier framework, however, Kumbhakar et al. (2009) specify a model where choices are binary (over production technologies) and where there is no spatial production component. Our specification involves a polychotomous choice (over potential worker teams) within peer-effect networks. Greene (2010) also has no spatial production component and assumes the selection bias works through the random noise, not inefficiency. Recently, Glass, Kenjegalievay, Sickles and Weyman-Jones (2016) and Glass, Kenjegalievay aand Sickles (2016) propose a spatial autoregressive (SAR) stochastic frontier model for panel data, but the weight matrix for the spatial dependency in their models is assumed to be exogenous. In the present framework the network is endogenous due to manager selectivity (and not worker selectivity), however, it is plausibly exogenous once we condition on the manager’s choices.

The rest of this paper is organized as follows. The next section introduces the econometric specification and the estimation approaches of our peer-effect stochastic frontier model with selectivity, and Section 3 provides an empirical exercise, using data from the National Basketball Association. Section 4 concludes.
1.2 Model and Estimation

1.2.1 Econometric model

Horrace, Liu and Patacchini (2016), hereafter HLP, consider a single firm where a manager has a choice set of worker groups at time $t$, $A_t = \{1,...,s,...\}$, that he can assign to a project, and she chooses a group of workers, $s$ in $A_t$ if only if,

$$d_{st}^* > \max_{j \neq s, j \in A_t} d_{jt}^*$$

where $d_{jt}^*$ is the manager’s selection equation, and it will be parameterized in terms of group-level data in the sequel. For now define $\epsilon_{st}^* = \max_{j \neq s, j \in A_t} d_{jt}^* - d_{st}^*$ with distribution $F_{st}^*$, so group $s$ is chosen if only if $\epsilon_{st}^* < 0$. We augment the HLP network production function with a one-sided inefficiency term. Let $u_{st}$ be an $N_t$-dimensional, worker-level inefficiency vector with typical element $u_{ist} > 0$, then our proposed model is:

$$Y_{st} = \lambda W_{st} Y_{st} + X_{1,st} \beta_1 + x_{2,st} \beta_2 1_{N_t} + v_{st} - u_{st},$$

where $Y_{st}$ is an $N_t$-dimensional output vector for the chosen group of workers, $\lambda$ is a scalar peer-effect parameter, $W_{st}$ is a weight matrix for the peer-effect (an $N_t \times N_t$ matrix with zeros on diagonal$^3$), $X_{1,st}$ is an $N_t \times k$ matrix of worker-varying variables, $x_{2,st}$ is a scalar worker-invariant variable, and $1_{N_t}$ is an $N_t \times 1$ vector of ones. The $v_{st}$ is an iid $N_t$-dimensional random vector representing statistical noise. Theoretical justification of the production function in equation (1.2) is presented in HLP, so we do not discuss it here. We make the following assumptions.

---

$^1$Horrace, Jung and Sanders, 2017, generalize the HLP model to multiple firms in direct competition.

$^2$If $Z_{jt}$ is random then $F_{st}^*$ is conditional on $Z_{jt}$.

$^3$A widely used example is the group averaging weight matrix, $W_{st}^* = \frac{1}{N_t - 1} (1_{N_t} 1_{N_t}^T - I_{N_t})$
Assumption 1.2.1 $W_{st}$, $X_{1,st}$ and $x_{2, st}$ are exogenous to $v_{st}$ and $u_{st}$.

Assumption 1.2.2 The elements of $v_{st}$ are distributed $iidN(0, \sigma_v^2)$, and the elements of $u_{st}$ are $iid$ exponentially distributed with probability density $g(u_{ist}) = \frac{1}{\sigma_u} e^{-u_{ist}/\sigma_u}$ and distribution function $G(u_{ist}) = 1 - e^{u_{ist}/\sigma_u}$. The error components are independent.

Assumption [1.2.1] is standard, and HLP provide conditioning arguments to establish exogeneity of the weight matrix. In essence, endogeneity of the weight matrix is not a result of unobserved worker behaviors; it is due to managerial selection, which can be corrected if the manager’s choices are observed. Assumption [1.2.2] follows Aigner, Lovell and Schmidt (1977) for the normal-exponential stochastic frontier model. While HLP allow statistical dependency between $\epsilon^*_{st}$ and $v_{st}$, we assume the following.

Assumption 1.2.3 The $\epsilon^*_{st}$ are independent of $v_{st}$ but dependent on $u_{ist}$ such that $\epsilon^*_{st}$ and $u_{ist}$ are jointly distributed $H(\epsilon^*_{st}, u_{ist}; \rho)$ with Pearson correlation $\rho$ and marginal distributions $G(u_{ist})$ and $F^*_{st}(\epsilon^*_{st})$. In our model managerial selection works through worker-level inefficiency. That is, the production function is stochastic through $v_{st}$, but the manager may know something about worker inefficiency, so he makes work group selections based on this knowledge. Following Dahl (2002), Bourguignon, Fournier and Gurgand (2007), and HLP, we need the following for identification of the parameters in equation (1.2).

Assumption 1.2.4 The joint distribution of $\epsilon^*_{st}$ and $u_{ist}$, $H(\epsilon^*_{st}, u_{ist}; \rho)$, does not depend on $Z_{jt}$.

Managerial selection of worker groups precludes identification of the parameters in the network production function of equation (1.2). HLP propose a two-step estimation procedure in the style of Heckman (1979) and Lee (1983), where in the first step the probability that a given group $s$ is selected in time $t$ is modeled as a function of a set of exogenous group-level
explanatory variables. In the second step, the estimates from the first step are used to construct a single index (e.g., an inverse Mill’s ratio in Heckman, 1979), which gets added to the right-hand side of equation (1.2) to adjust the specification for selection bias. HLP follow Lee (1983) and reduce the dimensionality of the selection bias with the scalar transformation $J_{st}(\cdot) \equiv F^{-1}(F^*_{st}(\cdot))$ of $\epsilon^*_{st}$, and we do the same. While HLP select the inverse of the standard normal for $F^{-1}$, we select the inverse of the exponential distribution with unit variance. Then, $J_{st}(\epsilon^*_{st})$ is an exponential random variable with unit variance by construction. For notational simplicity, define $\epsilon_{st} \equiv J_{st}(\epsilon^*_{st})$ and $\delta_{st} \equiv J_{st}(0)$, then the probability of group $s$ being selected in time $t$ is given by $P(\epsilon_{st} < \delta_{st} | \{Z_{jt}\}_{j \in A_t}) = F(\delta_{st})$, where $\delta_{st}$ is a $N_t$-dimension vector of exogenous variables that inform the manager’s selection decision.

With this transformation, $H(\epsilon^*_{st}, u_{ist}; \rho)$ could be specified as some bivariate exponential distribution, $BE(\epsilon_{st}, u_{ist}; \rho)$, whose marginal distributions are $F(\epsilon_{st})$ and $G(u_{ist})$. Frechet (1951) has shown that given two exponential distributions, there exist infinitely many bivariate exponential distributions with those given marginals. Hence, there exists a fundamental identification issue in constructing a bivariate exponential distribution from $F$ and $G$. However, using restrictions on their statistical independence, Gumbel (1960) proposes a Type I bivariate exponential family of distributions based on two exponential marginals that do not suffer from this issue. Therefore, we also assume the following.

**Assumption 1.2.5** The joint distribution of $\epsilon^*_{st}$ and $u_{ist}$, is a Gumbel Type I bivariate exponential:

$$H(\epsilon_{st}, u_{ist}; \rho(\alpha)) = 1 - e^{-\frac{u_{ist}}{\sigma_u}} - e^{-\epsilon_{st}} + e^{-\frac{u_{ist}}{\sigma_u} - \frac{u_{ist} \epsilon_{st} - \epsilon_{st}}{\sigma_u}} \quad \text{for } 0 \leq \alpha \leq 1.$$  

(1.3)

There are other classes of bivariate exponential distributions which possess this uniqueness property, and we leave exploration of these parametric alternatives to future research. One could also explore inducing cross-equation correlation with a copula.
Then the probability density is,

\[
h(\epsilon_{st}, u_{ist}; \rho(\alpha)) = \frac{1}{\sigma_u} e^{-\frac{\epsilon_{st}}{\sigma_u} - \frac{\alpha}{\sigma_u} u_{ist}\epsilon_{st} - \epsilon_{st}} \left[ (1 + \frac{\alpha}{\sigma_u} u_{ist})(1 + \alpha \epsilon_{st}) - \alpha \right], \tag{1.4}
\]

whose marginals are \(F\) and \(G\).

The Pearson correlation \(\rho\) is a monotonically deceasing function of \(\alpha\), which controls the level of dependence between \(\epsilon_{st}\) and \(u_{ist}\) with \(\alpha = 0\) corresponding to the independent case. Therefore, once we know \(\alpha\), then we know \(\rho(\alpha)\), uniquely. In particular, \(\rho(0) = 0\) and \(\rho(1) = -0.4837\). In other words, selecting this particular bivariate exponential requires that the Pearson correlation between \(\epsilon_{st}\) and \(u_{ist}\) be negative and not exceed 0.4837 in magnitude. While Assumption 1.2.5 restricts \(\alpha\) and \(\rho(\alpha)\), the negative \(\rho(\alpha)\) restricts the effect of the manager’s selection on output to be positive, as long a \(\alpha \neq 0\). That is, the expectation of worker-level inefficiency will decrease as the probability of a given group being selected increases. In other words, as the manager makes more effort to select a better group according to her choice rule, the outcome from the group will improve on average\footnote{A less restrictive bound is \(-1 \leq \rho \leq 0\), but to the best of our knowledge, there is no such bivariate exponential distribution with exponential marginals. Distributions with the bound of \(0 \leq \rho \leq 1\) exist but they may not be consistent with our model.}. This is not a feature of the Horrace, Liu and Patacchini (2016) model, where more managerial effort can lead to less output. This is an important contribution of our paper.

The effect of manager selection on worker-level inefficiency can be seen from the expectation of \(u\) conditional on the manger’s selection, which is,

\[
E(u_{ist} | \epsilon_{st} < \delta_{st}) = \frac{\sigma_u}{P_{st}} \left[ 1 - \frac{1}{\kappa_{st}} e^{-\delta_{st}} \right] \quad \text{where} \quad P_{st} = 1 - e^{-\delta_{st}} \quad \text{and} \quad \kappa_{st} = 1 + \alpha \delta_{st} \tag{1.5}
\]

See the derivation in Appendix A. From (1.5), we can verify that \(\lim_{\alpha \to 0} E(u_{ist} | \epsilon_{st} < \delta_{st}) = \sigma_u\) which implies that if there’s no correlation, this model reduces to the stochastic frontier model of Aigner et al (1977) with exponential one-sided error and peer-effects. We can also
Figure 1.1: (a) Bivariate exponential distribution of $u$ and $\epsilon$; (b) $E(u_{ist}|\epsilon_{st} < \delta_{st})$ by the value of $\alpha$ when $\sigma_{u} = 1$

show that $E(u_{ist}|\epsilon_{st} < \delta_{st})$ is a monotonic decreasing function of $\delta_{st}$ (if $\alpha \neq 0$) such that $\lim_{\delta_{st} \to \infty} E(u_{ist}|\epsilon_{st} < \delta_{st}) = \sigma_{u}$, which implies, if the probability of a given group being selected is higher, inefficiency is more likely to be small. Figure 1.1 (a) plots the bivariate exponential distribution of $u_{ist}$ and $\epsilon_{st}$ with $\sigma_{u} = 1$ and $\alpha = 0.7$, and Figure 1.1 (b) contains plots of $E(u_{ist}|\epsilon_{st} < \delta_{st})$ for various values of $\alpha$ when $\sigma_{u} = 1$.

1.2.2 Estimation of the Stochastic Frontier

Following Aigner et al (1977), the likelihood of the composed error, $\mu_{ist} = \nu_{ist} - u_{ist}$, conditional on $\epsilon_{st} < \delta_{st}$, is based on the "signal to noise" re-parameterization of the conditional density,

$$ h(\mu_{ist}|\epsilon_{st} < \delta_{st}) = \int_{0}^{\infty} h(u_{ist}, \mu_{ist}|\epsilon_{st} < \delta_{st}) du $$

$$ = \frac{1}{\sigma_{u} P_{st}} \left[ \Phi \left( -\frac{\mu_{ist}}{\sigma_{u}} - \frac{\mu_{ist} u_{ist}}{2\sigma_{u}^{2}} \right) e^{\frac{\mu_{ist} u_{ist}}{\sigma_{u}^{2}}} - \frac{\kappa_{st}}{\sigma_{u}} \Phi \left( -\frac{\mu_{ist}}{\sigma_{u}} - \frac{\kappa_{st} \mu_{ist}}{\sigma_{u}} \sigma_{v} \right) e^{\frac{\kappa_{st} \mu_{ist}}{\sigma_{u}^{2}} + \frac{\kappa_{st}^{2} \sigma_{v}^{2}}{2\sigma_{u}^{2}}} \right] $$

$$ = \frac{\eta}{\sigma_{v} P_{st}} \left[ \Phi \left( -\frac{\mu_{ist}}{\sigma_{v}} - \eta \kappa_{st} \right) e^{\frac{\mu_{ist} u_{ist}}{\sigma_{v}^{2}}} + \frac{\eta^{2}}{2} \frac{\kappa_{st} \mu_{ist}^{2}}{2\sigma_{v} \sigma_{u}} \right] $$

$$ \left( 1.6 \right) $
where \( \eta = \frac{\sigma_v}{\sigma_u} \). See the derivation in Appendix A. Then, the log-likelihood function is,

\[
\ln L = \sum_t \sum_{i \in s} \ln h(\epsilon_{st} < \delta_{st})h(\mu_{ist}|\epsilon_{st} < \delta_{st}) = \sum_t N_t \ln \frac{\eta}{\sigma_v} + \sum_t \ln |N_t - \lambda W_{st}|
\]

\[+
\sum_t \sum_{i \in s} \ln \left[ \Phi\left( -\frac{\mu_{ist}}{\sigma_v} - \eta \right) e^{-\frac{\mu_{ist}^2}{2\sigma_v}} - \frac{\kappa_{st}}{\sigma_v} \Phi\left( -\frac{\mu_{ist}}{\sigma_v} - \kappa_{st}\eta \right) e^{-\frac{\mu_{ist}\kappa_{st}}{\sigma_v} + \frac{\kappa_{st}^2}{2}} \right] \]  

(1.7)

where \( w_{st,ij} \) is the \( ij^{th} \) element in \( W_{st} \). \( |N_t - \lambda W_{st}| \) is the determinant of the Jacobian of the transformation from \( \mu_{ist} \) to \( y_{ist} \), which is undertaken in a standard spatial model to account the endogeneity of the spatial lag of the dependent variable (Anselin, 1988; Elhorst, 2010).

To make the estimation feasible we substitute and \( \hat{\mu}_{ist} = y_{ist} - \lambda \sum_j w_{st,ij} y_{jst} - x_{st}\beta \) for \( \mu_{ist} \) in the likelihood and maximize it. There are a number of ways to implement the maximum likelihood estimation of (1.7) but, to simplify the estimation, we consider a two-step approach here and provide another possible three-step approach in Appendix A.

### 1.2.3 A Two-Step Approach

Our two-step estimation approach is as follows.

**Step 1:** With specification (1.1), if we assume \( \xi_{st} \) is independently and identically Gumbel distributed, then \( \gamma \) can be estimated by maximizing the log-likelihood below (McFadden, 1974),

\[
\ln L_1 = \sum_t \left( Z_{st} \gamma - \ln \sum_{j \in A_t} e^{Z_{jt}\gamma} \right) + C
\]

(1.8)

where \( C \) is the constant which does not depend on the \( \gamma \). From this conditional logit estimator, \( \hat{\gamma} \), we can calculate the probability of the group \( s \) being chosen as \( \hat{P}_{st} = \frac{e^{Z_{st}\hat{\gamma}}}{\sum_{j \in A_t} e^{Z_{jt}\hat{\gamma}}} \). Then we can compute \( \hat{\delta}_{st} = F^{-1}(\hat{P}_{st}) \) where \( F^{-1} \) is the inverse of the exponential distribution with unit variance, which will be \( F^{-1}(x) = -\ln(1 - x) \).
Step 2: Maximize the log-likelihood in (1.7) to get $\hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2, \hat{\eta}, \hat{\sigma}_v$, and $\hat{\alpha}$. Because we use multiple step estimation, we use the Murphy and Topel (2002) correction to adjust standard errors. We provide relevant derivatives of the log-likelihood for the production function and selection equation in Appendix A.

1.2.4 Estimation of Individual Inefficiencies

Next, we can estimate worker inefficiency as $E(u_{ist}|\mu_{ist}, \epsilon_{st} < \delta_{st})$, following Jondrow et al. (1982). That is, the conditional mean function is given by,

$$E(u_{ist}|\mu_{ist}, \epsilon_{st} < \delta_{st}) = \int_0^\infty u \frac{h(u_{ist}, \mu_{ist}, \epsilon_{st} < \delta_{st})}{h(\mu_{ist}, \epsilon_{st} < \delta_{st})} du = \int_0^\infty u \frac{h(u_{ist}, \mu_{ist}|\epsilon_{st} < \delta_{st})}{h(\mu_{ist}|\epsilon_{st} < \delta_{st})} du$$

$$= \frac{A_{ist}}{A_{ist} - B_{ist}} \left( \Lambda_{ist} + \sigma_v \left[ \frac{\phi(-\Lambda_{ist}/\sigma_v)}{\Phi(\Lambda_{ist}/\sigma_v)} \right] \right) - \frac{B_{ist}}{A_{ist} - B_{ist}} \left( \Lambda^*_{ist} + \sigma_v \left[ \frac{\phi(-\Lambda^*_{ist}/\sigma_v)}{\Phi(\Lambda^*_{ist}/\sigma_v)} \right] \right)$$

where $A_{ist} = \Phi(\Lambda_{ist}/\sigma_v) e^{\mu_{ist}/\sigma_u} + \frac{\sigma^2_v}{2\sigma_u^2}$, $B_{ist} = \frac{\kappa_{ist}}{\sigma_u} \Phi(\Lambda^*_{ist}/\sigma_v) e^{\frac{\kappa_{ist}\mu_{ist}}{\sigma_u} + \frac{\kappa_{ist}^2\sigma^2_v}{2\sigma_u^2}}$,

$$\Lambda_{ist} = -\mu_{ist} - \frac{\sigma^2_v}{\sigma_u} \quad \text{and} \quad \Lambda^*_{ist} = -\mu_{ist} - \frac{\kappa_{ist}\sigma^2_v}{\sigma_u} \quad (1.9)$$

See the derivation in Appendix A. As previously stated, as $\alpha$ goes to zero or $\delta_{st}$ goes to infinity, this conditional expectation goes to $\sigma_u$, implying that there is no selection bias induced by the manager, and her choices have no effect on worker performance. Estimation proceeds by plugging likelihood estimates of the variance components into equation (1.9).

---

6In order to reduce the number of parameters we need to estimate, we may replace the $\beta_1$ with the consistent estimator of $\beta_1$ conditional on $\lambda$ which is $\hat{\beta}_1 = (X_1'QX_1)^{-1}X_1'Q(Y - \lambda WY)$, where $Q$ is the within transformation matrix, and $X_1$ and $Y$ are the matrices containing all the observations over $t$.\[486,110]
1.3 Empirical Application

1.3.1 Data and Variables

We apply our network stochastic frontier model to NBA data for three teams over the 82-game 2015-2016 regular season: the Detroit Pistons (DET), the Denver Nuggets (DEN) and the Atlanta Hawks (ATL). The data were purchased and downloaded from BigDataBall.com. The three teams are from different divisions (Central, Northwest and Southeast divisions, respectively). In recent years, they have all been ranked at or near the median NBA team in terms of their Rating Percentage Index (RPI) which measures the overall strength of teams based on win percentage, opponent win percentage, and win percentage of opponents’ opponents. Our data and specification are designed to mimic HLP analysis of the Syracuse Men’s Basketball team, but for a few exceptions.

Following HLP, we define a time period, \( t \), in a given game as the interval between two consecutive substitutions by either team. With this definition our data initially contain 1,450 time periods for Detroit, 1,520 periods for Denver and 2,068 periods for Atlanta. Over an 82 regular game NBA season, this is an average period duration of 2.82, 2.76 and 2.15 minutes per period, respectively. Compare these averages to 2.46 minutes per period for the 2011-12 Syracuse University Men’s Basketball team (and their opponents) as analyzed in HLP, and it seems that NBA coaches make substitutions as frequently as NCAA coaches. Following HLP, we drop time periods that are less than 30 seconds in length, and also drop overtime periods. After dropping the periods we are left with 1,165 time periods for DET, 1,205 time periods for DEN, and 1,571 time periods for ATL over the course of the season.\(^8\)

\(^7\)The RPI for NBA is obtained from ESPN: http://www.espn.com/nba/stats/rpi/year/2016. Our purpose is to demonstrate the empirical importance of the proposed model, so we do not estimate a model for every NBA team. See Horrace, Jung and Sanders (2017, working paper) for a league-wide analysis of these data.

\(^8\)HLP also drop the time periods if the number of player types in any given active lineup of the periods is less than 2 where “types” are the player types such as Guards or Forwards. We found that this would remove a significant number of time periods in our case, so we did not drop those periods in our analysis.
players per team on the court at all times this leads to 5,825, 6,025 and 7,845 observations, respectively.

We use the "same-type" peer-effect weight matrix considered in HLP, where "types" are the player types: *Guards* or *Forwards*, with *Forwards* including centers. That is, the same-type weight matrix is $W$, where $W_0 = [w_{0,ij}]$ is an adjacency matrix with $w_{0,ij} = 1$ if the $i^{th}$ and $j^{th}$ players are both guards or forwards. Then row-normalize $W_0$ so that $W_1N = 1_N$. This network specification assumes that each individual is affected only by the same type of agents in his network (exclusion restrictions). These exclusion restrictions allow us to use the 3-step estimation procedure discussed in Appendix A. For details about the relationship between topology of weight matrix and the estimation procedure, see Appendix A.

For example, let’s assume the lineup for DET is $[F, F, G, F, G]'$ where $F = forward$ and $G = guard$. Then, the same-type peer-effect weight matrix is given by:

$$W = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

We use two output measures that are slightly different than those in HLP. The first is NBA player efficiency based on the work of sports economist, David Berri: $y_{ist} = (0.064 \times 3PT_{ist} + 0.032 \times 2PT_{ist} + 0.017 \times FT_{ist} + 0.034 \times REB_{ist} + 0.033 \times STL_{ist} + 0.020 \times BLK_{ist} - 0.034 \times MFG_{ist} - 0.015 \times MFT_{ist} - 0.034 \times TO_{ist})/Mins_{ist}$, where $3PT_{ist}$, $2PT_{ist}$ and $FT_{ist}$ are 3-point field goals made, 2-points field goals made, and free throws made (respectively), $REB_{ist}$ is rebounds, $AST_{ist}$ is assists, $STL_{ist}$ is steals, $BLK_{ist}$ is blocked shots, $MFG_{ist}$ is missed field goals, $MFT_{ist}$ is missed free throws, $TO_{ist}$ is turn overs, and $Mins_{ist}$ is minutes played by player $i$ in line-up $s$ in period $t$. This statistics is one of the more popular metrics used

---

Footnote: HLP use both same- and cross-type weight matrices. The different-type weight matrix is $W^d$, where $W^d_0 = [w^d_{0,ij}]$ is an adjacency matrix with $w^d_{0,ij} = 1$ if the $i^{th}$ and $j^{th}$ players are a guard and a forward (or vice versa). Then row-normalize $W^d_0$ so that $W^d_1N = 1_N$. However, they noticed that including both weight matrices into the model at the same time may lead to a multicollinearity problem. Therefore, we only include the same type weight matrix in this exercise.
to measure the value of an NBA player’s production. HLP use a similar metric, but the weights assigned to each element of their statistic are different. Unfortunately, both the HLP and Berri efficiency measures can frequently produce \( y_{ist} = 0 \), when the duration of \( t \) is small. That is, not every player can be productive in every 30-second interval, and many of these zeros can lead to likelihood function instability. Therefore, we also experiment with a cumulative version of player efficiency, measured from the start of the game, instead of in an individual time period. This cumulative measure smooths the variability of productivity when an active player moves from a highly productive period to a non-productive period.

The rest of the variables in the analysis and the production function specification are identical to HLP, and this is by design. The player-varying exogenous variables in the outcome equation (\( X_{1,ist} \)) are \( Experience_{ist} \) and \( Fatigue_{ist} \). \( Experience_{ist} \) is minutes played from the start of the game to the end of period \( t-1 \), and \( Fatigue_{ist} \) is minutes continuously played until the end of period \( t-1 \). The player-invariant exogenous variables (\( x_{2,tsi} \)) are the RPI of the opposing team from the end of the previous season, \( Home \), an indicator variable for a home game, and \( 2ndHalf \), an indicator variable equal to 1 if \( t \) is in the second half.

The exogenous variables in the selection equation (\( Z_{st} \)) are lineup-level aggregations of variables from the production equation. \( Lineup – Efficiency_{st} \) is the sum of the \( y_{ist} \) of lineup \( s \) from the start of the game until the end of period \( t-1 \). \( Lineup – Experience_{st} \) is the sum of minutes played by the lineup at the end of period \( t-1 \). \( Lineup – Fatigue_{st} \) is the total minutes continuously played by the lineup at the end of period \( t-1 \). \( Lineup – Fouls_{st} \) is the total fouls by the lineup at the end of period \( t-1 \). \( One – Substitution_{st} \) is a dummy equal to 1, if one player was substituted to achieve the current lineup. \( Two – Substitution_{st} \) is a dummy equal to 1, if two players were substituted to achieve the lineup at time \( t \). We included player dummies to control for player-specific heterogeneity.

The descriptive statistics for the continuous variables are presented in Table 1.1. The first two

\[ \text{See http://wagesofwins.com/how-to-calculate-wins-produced/}. \]
Table 1.1: Descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Team</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency (y)</td>
<td>Weighted sum of player productivity divided by minutes in period t</td>
<td>DET</td>
<td>0.0052</td>
<td>0.0222</td>
<td>0.1818</td>
<td>-0.1200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>0.0052</td>
<td>0.0221</td>
<td>0.2226</td>
<td>-0.1133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>0.0059</td>
<td>0.0239</td>
<td>0.1800</td>
<td>-0.1749</td>
</tr>
<tr>
<td>Cumulative efficiency (y)</td>
<td>Average efficiency until the end of period t</td>
<td>DET</td>
<td>0.0057</td>
<td>0.0097</td>
<td>0.0949</td>
<td>-0.0618</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>0.0057</td>
<td>0.0101</td>
<td>0.1200</td>
<td>-0.0638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>0.0062</td>
<td>0.0101</td>
<td>0.1042</td>
<td>-0.0618</td>
</tr>
<tr>
<td>Experience (X₁)</td>
<td>Minutes played from the start of the game to the end of period t-1</td>
<td>DET</td>
<td>14.1664</td>
<td>9.2997</td>
<td>43.1833</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>13.060</td>
<td>8.7835</td>
<td>40.933</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>12.954</td>
<td>8.6636</td>
<td>38.833</td>
<td>0</td>
</tr>
<tr>
<td>Fatigue (X₁)</td>
<td>Minutes continuously played until the end of period t-1</td>
<td>DET</td>
<td>4.1848</td>
<td>4.5555</td>
<td>31.183</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>4.4203</td>
<td>4.6919</td>
<td>40.450</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>3.0356</td>
<td>3.2668</td>
<td>19.133</td>
<td>0</td>
</tr>
<tr>
<td>RPI (x₂)</td>
<td>Previous year RPI of the opposing team.</td>
<td>DET</td>
<td>0.4989</td>
<td>0.0414</td>
<td>0.597</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>0.5033</td>
<td>0.0437</td>
<td>0.597</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>0.4999</td>
<td>0.0380</td>
<td>0.597</td>
<td>0.405</td>
</tr>
<tr>
<td>Lineup efficiency (Z)</td>
<td>Total efficiency of the players in the lineup until the end of period t-1</td>
<td>DET</td>
<td>0.0274</td>
<td>0.0224</td>
<td>0.2993</td>
<td>-0.0661</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>0.0270</td>
<td>0.0236</td>
<td>0.3161</td>
<td>-0.1602</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>0.0291</td>
<td>0.0225</td>
<td>0.1805</td>
<td>-0.1073</td>
</tr>
<tr>
<td>Lineup experience (Z)</td>
<td>Total minutes played by the players in the lineup until the end of period t-1</td>
<td>DET</td>
<td>70.832</td>
<td>38.358</td>
<td>181.38</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>65.302</td>
<td>37.288</td>
<td>166.50</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>64.774</td>
<td>37.724</td>
<td>175.20</td>
<td>4.7667</td>
</tr>
<tr>
<td>Lineup fatigue (Z)</td>
<td>Total minutes continuously played by the players in the lineup until the end of period t-1</td>
<td>DET</td>
<td>20.924</td>
<td>11.545</td>
<td>64.633</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>22.101</td>
<td>10.692</td>
<td>67.116</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>15.178</td>
<td>7.5981</td>
<td>50.266</td>
<td>0</td>
</tr>
<tr>
<td>Lineup fouls (Z)</td>
<td>The total number of fouls of the players in the lineup until the end of period t-1</td>
<td>DET</td>
<td>4.7974</td>
<td>3.4686</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DEN</td>
<td>4.5710</td>
<td>3.4011</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATL</td>
<td>4.4799</td>
<td>3.3696</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>
columns of the table contains the variable name and a brief definition, receptively. The rest of the table presents the means, standard deviations and extrema for each team and variable. For example, the average player Efficiency is 0.0052, 0.0052 and 0.0059 for DET, DEN and ATL, respectively. Not surprisingly, at the end of the season ATL had the best win percentage among the three teams (ATL: 0.585, DET: 0.537, and DEN 0.402). Therefore, average player efficiency seems to be positively correlated with wins in this sample. These results are true of the Cumulative Efficiency measure too, with cumulative efficiency being higher on average than the average of the per period marginal efficiency (Efficiency). Compare 0.0052 to 0.0057 (DET), 0.0052 to 0.0057 (DEN) and 0.0059 to 0.0062 (ATL). We can also see that the standard deviation is everywhere lower for the cumulative measure, as expected. Both the Experience and Fatigue variables are lowest on average for Atlanta, implying that the Atlanta coach tends to make substitutions more frequently than both Detroit and Denver. Strength of schedule as measured by RPI seems to be about the same across the three teams, but remember that this is a static variable measured once at the end of last season for each team. The last 4 variables in Table 1.1 are "per lineup" averages across the different lineups used by each team. Since there are exactly five active players in any period, the lineup averages tend to be about 5 times larger than the “per player" averages. This is certainly the case when we compare Efficiency (in the first row) to Lineup Efficiency (sixth row). This also appears to be the case, if we compare experience and lineup experience or if we compare fatigue and lineup fatigue for each team. Finally, the average number of fouls across lineups is between 4 and 5 for all three teams.

1.3.2 Estimation Results

Estimation results for the selection model are in Table 1.2 and the results are sensible. For example, efficient lineups (the first variable in the table) have a higher probability of being selected by the coaches, while lineups with larger number of fouls (Lineup Fouls)
and longer continuous play without rest (Lineup Fatigue) are less likely to be selected. In case of the Lineup Experience variable, the Detroit coach picks experienced lineups with higher probability (coefficients of 0.0135) but the Denver and the Atlanta coefficients are not significant. Overall, the first stage selection results are compelling. We now turn to our main results in Table 1.3.

Table 1.2: Estimation of the selection equation

<table>
<thead>
<tr>
<th>Dep. var.: probability of lineup selection</th>
<th>DET</th>
<th>DEN</th>
<th>ATL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lineup efficiency</td>
<td>6.8142**</td>
<td>6.1258**</td>
<td>5.7481***</td>
</tr>
<tr>
<td>(Tstat)</td>
<td>(2.560)</td>
<td>(2.233)</td>
<td>(2.661)</td>
</tr>
<tr>
<td>Lineup fouls</td>
<td>-0.1312***</td>
<td>-0.2330***</td>
<td>-0.0509**</td>
</tr>
<tr>
<td></td>
<td>(-5.068)</td>
<td>(-8.521)</td>
<td>(-2.316)</td>
</tr>
<tr>
<td>Lineup experience</td>
<td>0.0135***</td>
<td>-0.0061</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td>(4.199)</td>
<td>(-1.517)</td>
<td>(-0.371)</td>
</tr>
<tr>
<td>Lineup fatigue</td>
<td>-0.1345***</td>
<td>-0.1460***</td>
<td>-0.1803***</td>
</tr>
<tr>
<td></td>
<td>(-19.444)</td>
<td>(-19.619)</td>
<td>(-22.896)</td>
</tr>
<tr>
<td>One Substitution</td>
<td>5.2378***</td>
<td>7.1250***</td>
<td>6.1627***</td>
</tr>
<tr>
<td></td>
<td>(28.603)</td>
<td>(27.957)</td>
<td>(35.703)</td>
</tr>
<tr>
<td>Two Substitution</td>
<td>2.7571***</td>
<td>4.0893***</td>
<td>3.5791***</td>
</tr>
<tr>
<td></td>
<td>(18.295)</td>
<td>(19.184)</td>
<td>(24.907)</td>
</tr>
</tbody>
</table>

Log-likelihood: -1,882.78, -1,973.06, -2,938.02
Sample size: 1,165, 1,205, 1,571

T statistic in parentheses
Statistical significance: *** p < 0.01; **p<0.05; *p<0.1.

After a few preliminary estimations, we discovered that $\eta = \sigma_v/\sigma_u$ is large in these data (sometimes larger than 10). In other words, there is little player inefficiency on these teams. In this case, the likelihood of (1.7) has trouble converging and is quite sensitive to outliers. This makes sense because our model is basically shrinking the inefficiency estimate, $\sigma_u$, and attributing part of it to a managerial selection effect, $\alpha$. If inefficiency is too small in data, then there’s nothing to shrink so the correlation parameter can’t be well-identified, which is manifest as instability in the likelihood.

To minimize instability, we use the 3-step estimation procedure discussed in Appendix A.
After estimating the selection equation, we use Conditional Maximum Likelihood (CML) estimation (Lee, 2007) to estimate \( \lambda \) and \( \beta_1 \) and then use the likelihood to estimate the rest of the parameters in the outcome equation. In this procedure, we calculate the standard errors for the parameters from the analytic forms in Appendix A.

Table 1.3: Three step ML estimation of the outcome function

<table>
<thead>
<tr>
<th></th>
<th>DET</th>
<th></th>
<th>DET</th>
<th></th>
<th>ATL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Peer-effects</td>
<td>0.0441***</td>
<td>0.0297**</td>
<td>-0.0219</td>
<td>-0.0011</td>
<td>0.0261**</td>
<td>0.0200**</td>
</tr>
<tr>
<td>(Tstat)</td>
<td>(3.328)</td>
<td>(2.316)</td>
<td>(-0.465)</td>
<td>(-0.096)</td>
<td>(2.460)</td>
<td>(2.332)</td>
</tr>
<tr>
<td>Experience</td>
<td>-0.00002</td>
<td>0.00007***</td>
<td>0.00007</td>
<td>0.00014***</td>
<td>0.00022***</td>
<td>0.00022***</td>
</tr>
<tr>
<td></td>
<td>(-0.099)</td>
<td>(2.395)</td>
<td>(0.083)</td>
<td>(4.448)</td>
<td>(3.237)</td>
<td>(7.472)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.00017***</td>
<td>0.00014***</td>
<td>0.00005</td>
<td>0.0007*</td>
<td>-0.00002</td>
<td>0.00006</td>
</tr>
<tr>
<td></td>
<td>(2.920)</td>
<td>(3.956)</td>
<td>(0.099)</td>
<td>(1.712)</td>
<td>(-0.267)</td>
<td>(1.329)</td>
</tr>
<tr>
<td>RPI</td>
<td>-0.0089</td>
<td>-0.0039</td>
<td>-0.0096</td>
<td>-0.0120***</td>
<td>-0.0337***</td>
<td>-0.0185***</td>
</tr>
<tr>
<td></td>
<td>(-0.495)</td>
<td>(-1.365)</td>
<td>(-0.222)</td>
<td>(-4.099)</td>
<td>(-4.464)</td>
<td>(-6.084)</td>
</tr>
<tr>
<td>Home</td>
<td>0.0014</td>
<td>0.0012***</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0005**</td>
</tr>
<tr>
<td></td>
<td>(1.611)</td>
<td>(4.589)</td>
<td>(0.089)</td>
<td>(0.451)</td>
<td>(0.444)</td>
<td>(2.247)</td>
</tr>
<tr>
<td>2nd Half</td>
<td>-0.0002</td>
<td>-0.0008**</td>
<td>-0.0008</td>
<td>-0.0018***</td>
<td>-0.0022**</td>
<td>-0.0026***</td>
</tr>
<tr>
<td></td>
<td>(-0.063)</td>
<td>(-2.124)</td>
<td>(-0.056)</td>
<td>(-4.232)</td>
<td>(-2.214)</td>
<td>(-6.828)</td>
</tr>
<tr>
<td>( \eta = \frac{\sigma_\nu}{\sigma_v} )</td>
<td>16.2392</td>
<td>2.3223***</td>
<td>7.1804**</td>
<td>2.2514***</td>
<td>3.1830***</td>
<td>2.9764***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(7.197)</td>
<td>(2.112)</td>
<td>(11.149)</td>
<td>(4.586)</td>
<td>(4.817)</td>
</tr>
<tr>
<td>( \sigma_\nu )</td>
<td>0.0217</td>
<td>0.0080</td>
<td>0.0214</td>
<td>0.0082</td>
<td>0.0225</td>
<td>0.0090</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.3623</td>
<td>0.0000</td>
<td>0.4691</td>
<td>0.0006</td>
<td>0.0083</td>
<td>0.0054</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.000)</td>
<td>(0.795)</td>
<td>(0.006)</td>
<td>(0.028)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Player dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>4,168.56</td>
<td>9,509.94</td>
<td>5,127.80</td>
<td>10,412.96</td>
<td>3,566.96</td>
<td>10,721.90</td>
</tr>
<tr>
<td>Sample size</td>
<td>5,825</td>
<td>5,825</td>
<td>6,025</td>
<td>6,025</td>
<td>7,845</td>
<td>7,845</td>
</tr>
</tbody>
</table>

Model 1: \( y = \text{efficiency} \), Model 2: \( y = \text{cumulative efficiency} \)
T statistics in parentheses, and Statistical significance: *** \( p < 0.01 \); ** \( p<0.05 \); * \( p<0.1 \).

The three-step ML estimation results for the outcome equation are in Table 1.3. Model 1 corresponds to the dependent variable in the outcome equation being "per period" efficiency, while Model 2 is based on cumulative efficiency. The results may be summarized as follows.

PEER-EFFECT: Overall, the peer-effects for these teams are smaller than those reported

---

11In calculating the Hessian, we removed observations which produced explosive values in the Hessian. In Model 1 for DET and DEN, we removed four and three observations, respectively, in this calculation. We always applied the Murphy and Topel correction to our standard errors.
in HLP for the Syracuse Men’s basketball team. Also, there is some heterogeneity in the
magnitude and the direction of the peer-effects across teams. In particular, DET and ATL
show significant and positive peer-effects in both model 1 and 2 while DEN shows a negative
but insignificant peer-effect in all cases\textsuperscript{12} The insignificant results of DEN may be due to a
poorly specified model or it may truly tell us something about the nature of the different
teams. In particular the peer effects may be measures of “team chemistry,” all things being
equal.

INEFFICIENCY: When we use cumulative efficiency for the outcome measure (Model 2), η
is significant for all three teams, which may be due to the fact that the cumulative measure
has lower overall variance, but the unexplained portion of the variance seems to be moved
into the inefficiency term, increasing the signal to noise ratio (decreasing η). For DET, Model
1 produces an insignificant η equal to 16.2392, but in Model 2 it is 2.3223 and significant.
In general the η estimate is relatively large across teams and models, implying that there
appears to be little player inefficiency in the data. When this turns out to be true, HLP may
be the preferred specification\textsuperscript{13}

COACHING EFFECT: In all cases, selection has no effect on outcomes. That is, α is always
insignificant. First, the production function and the selection equation used in this application
may simply be incorrect. The data may require a more sophisticated modeling approach
or richer data. Second, the optimization rule for coaching may not be to just minimize
instantaneous player inefficiency. Coaches may behave more strategically than our simple
model may allow. Third, we conducted a brief simulation study and found that is some cases
α is only weakly identified, which may explain the very small values of the estimates. When
there is no selection effect on the outcome equation, our model reduces to the exponential

\textsuperscript{12} A sufficient condition for stationarity is that the peer-effects should be on the unit circle.
\textsuperscript{13} As we used the 3 step approach with CML for the second step, the results for the first and the second
step between HLP and our model are the same. The estimation results using HLP model on this data can be
obtained by request.
normal model of Aigner et al. (1977).\footnote{Using \( \hat{\lambda} \) and \( \hat{\beta}_3 \) from CML, we estimated the ALS-EN for \( \beta_2, \eta, \) and \( \sigma_v \) and found that the estimates are (almost) identical between the two models. The standard errors for those parameters calculated from the two models are not that different except for \( \eta, \) which is due to the Murphy and Topel correction employed on the variance-covariance matrix of our model (The standard error for \( \eta \) from our model is two or three times bigger than the one from ALS-EN).}

1.4 Conclusions

We consider a stochastic production function model that extends the basic normal-exponential model of Aigner et al. (1977) to incorporate worker peer-effect networks and managerial selection that works through a time-varying worker inefficiency term. In this setting we have allowed for correlation between the outcome and selection equations, using a Gumbel Type I bivariate exponential distribution for the inefficiency component of the outcome equation and the selection equation error. Our model is an improvement over the HLP model in the sense that the one-sided selection error restricts the manager to worker networks that will only decrease worker inefficiency and improve individual worker productivity. Managerial selections can only increase output in our model, which is a theoretically appealing restriction. One downside of the model is that the Pearson correlation between the outcome and the selection equation cannot be too large, otherwise the model is not identified. However, this is the cost we incur for adding a third variance parameter, \( \alpha, \) to the already complicated stochastic frontier model. In addition to presenting the theoretical model we detailed two estimation approaches that are similar in spirit to those of HLP.

We apply the model to estimate peer-effects for three NBA teams from the 2015-16 regular season. Our results were mixed, with some instability in the likelihood function. Nonetheless some of the results, particularly on the peer-effects, are compelling. In our example, we limited ourselves to the production function specification and network topology for college basketball considered in HLP, except we selected a sightly different measure for player efficiency. This may partly explain why our results were unstable: the NBA is a different game than the...
NCAA, so perhaps future work should consider other specifications of and topologies in the production function. Both the HLP model and the present model treat the strategy of the opposing coach as exogenous (as measured by the opposing teams exogenous RPI). In most network industries it may be reasonable to suspect the productive behavior of competitors to be exogenous, but certainly not in sports where strategic reactions are commonplace. Developing network production models that endogenize the strategy of the opposing team would be an important contribution for researchers interested in the productivity of athletes in competitive team sports. This is currently being explored by the authors.
References


Network Competition and Team Chemistry in the NBA

We consider a heterogeneous social interaction model where agents interact with peers within their own network but also interact with agents across other (non-peer) networks. To address potential endogeneity in the networks, we assume that each network has a central planner who makes strategic network decisions based on observable and unobservable characteristics of the agents in her charge. The model forms a simultaneous equation system that can be estimated by Quasi-Maximum Likelihood or Generalized Method of Moments. We apply a restricted version of our model to data on NBA games, where agents are players, networks are individual teams organized by coaches, and at any time a player only interacts with two networks: their team and the opposing team. We find significant positive peer-effects (team chemistry) in NBA games.
2.1 Introduction

We consider a world with \( R \) independent networks where agents interact with peers within their own network but also interact with non-peers from other networks, but in different ways. For example, we can think of teams of individual agents that cooperate within their network but compete across networks. Competition between two or more firm R&D alliances comes to mind. A given firm may cooperate with an R&D ally to achieve an intellectual property discovery, but firms across alliances compete. Airline alliances (e.g., SkyTeam, Star Alliance and OneWorld) cooperate within their networks but compete across networks. In these examples, multiple networks or teams may simultaneously compete, but in some instances, such as sports, team competition is head-to-head. We restrict attention to models where an agent’s single outcome (e.g., sales performance) is a function of the simultaneous outcomes of their peers and competitors. In particular, we are not concerned with the case of Liu (2014) or Cohen-Cole et al. (2017), where there is a single peer network (no competitors) with multiple outcome variables (e.g., a single network of friends each of whom allocates effort to simultaneous outcomes, such as labor and leisure hours). In the aforementioned examples, within-network interaction is complementary and cross-network interaction is competitive; however, our model allows for the converse to be true. For example, in sports competition a team’s performance may be worsened or enhanced when they face a better team.

In these examples, social interaction decisions are likely to be guided by a central planner for each network (e.g., a sales manager), and the choices of the planner may induce what Manski (1993) calls a correlated effect, where “individuals in the same group tend to behave similarly because they... face similar institutional environments.” The usual solution to the correlated effects problem is to include a network-level fixed or random effect in the model specification. However, if the network consists of labor inputs to a production process, then the network itself may be endogenous in the same way that any production input may be
simultaneously (and strategically) selected with output in a manager’s profit maximization problem (Olley and Pakes, 1996). Following Horrace, Liu, and Patacchini (2016), we augment the outcome equation with a selection equation that models the decisions of the central planners’ network choices. We consider both parametric (Lee, 1983) and semi-parametric (Dahl, 2002) approaches to the selection problem. Horrace, Liu, and Patacchini (2016) consider a network production function where a manager selects workers into a network to produce output, but they ignore cross-network competition. In this sense, our paper is a generalization of their study.

Social network interactions\footnote{See Manski (1993), Moffitt (2001), Lee (2007a), and Bramoulle et al (2009)} have been studied extensively in recent decades; however, simultaneous cross-network interactions remain relatively unexplored. A few papers model simultaneous activity for a single network, and are multivariate extensions of the single equation Spatial Auto-Regressive (SAR) model of Cliff and Ord (1973, 1981) to simultaneously determined outcome variables. For example, Kelejian and Prucha (2004) generalize the SAR model to a simultaneous system. Baltagi and Deng (2015) extend the model to a panel setting with random effects, while Cohen et al. (2017) extend it to a simultaneous system with fixed effects. Yang and Lee (2017) study identification and Quasi-Maximum Likelihood (QLM) estimation of the model of Kelejian and Prucha (2004). Empirical implementation of these types of simultaneous models include the effect of peer networks on migration and housing prices (Jeanty et al., 2010); on migration, employment and income (Gebremariam et al., 2011); on rents for studio, one-bedroom and two-bedroom apartments (Baltagi and Bresson, 2011); on simultaneous fiscal policies (Allers and Elhorst, 2011); among others. All these models are clearly related, but they don’t consider multiple peer networks that may be engaged in simultaneous competition around a single outcome variable.

Aside from the applied econometric contributions mentioned above, another contribution of the study is empirical in nature. We apply our model to the 2013-14, 2014-15, and 2015-16
NBA regular seasons to estimate within-network and cross-network peer-effects for each team in the league. Team chemistry (team peer-effects) receives substantial attention as a factor influencing team performance in business and sports. Unfortunately, the concept is notoriously difficult to measure. Schrage (2014) discusses team chemistry measurement as “the new holy grail of performance analytics” in sport and business. McCaffery and Bickart (2013) estimate team chemistry as a function of biological synchrony among players, and Kraus, Huang, and Keltner (2010) find evidence that early-season, on-court tactile communication is a predictor of later-season success at both the individual and team levels. Horrace, Liu, and Patacchini (2016) develop a network production function model that estimates peer (on-court teammate) network effects upon player productivity in men’s college basketball. Due to data limitations, their peer-effects do not condition upon (confounding) effects from the network of on-court competitors. The absence of competitor effects in their model introduces omitted variable bias, and their peer-effects do not constitute (ceteris paribus) estimates of team-level chemistry.

We extend their peer-effect model to account for the strategic decisions and contemporaneous play of the opposing team by augmenting it with a competitor network, leading to estimation of a team’s “competitor-effect” in addition to its peer-effect. The interpretation of the team-level peer-effect is the same, but, in our model, the effect conditions on both teams’ strategies and abilities, making for a more reliable team-chemistry measure. We find that peer-effects are generally positive for NBA teams and are moderately persistent across team-season but that negative peer-effects occur.

The rest of this paper is organized as follows. The next section introduces the econometric specification and the estimation approaches, Section 3 provides the result of the empirical exercise and Section 4 concludes.
2.2 Model and estimation

2.2.1 Econometric model

Outcome function: We have $R$ networks (alliances, chains or teams), and each network, $r = 1, \ldots, R$ contains $n_r$ peers with $N = \sum_r n_r$. Peers cooperate within their own network but compete with members of the other networks. The time period, $t$, is suppressed here. In the model that follows, all the data, weighting matrices, and the error term vary with time, and all the parameters do not. The outcome function for the $r^{th}$ network is:

$$y_r = \lambda_{rr} W_{rr} y_r + \sum_{k \neq i} \lambda_{rk} W_{rk} y_k + x_r \beta_r + u_r, r = 1, \ldots, R,$$

(2.1)

where $y_r$ is an $n_r \times 1$ outcome vector for the $r^{th}$ network, $x_r$ is a $n_r \times p$ exogenous input matrix, and $u_r$ is an $n_r \times 1$ disturbance vector. $W_{rr}$ is an $n_r \times n_r$ weight matrix for interaction within the $r^{th}$ network, while $W_{rk}$ for $k \neq r$ is a $n_r \times n_k$ matrix for the effect from the $k^{th}$ network to the $r^{th}$ network. We assume the matrices have network structure and are row-normalized, so that $\lambda_{rr}$ is the average within-network effect for the $r^{th}$ network, and $\lambda_{rk}$ is the average cross-network effect for $k \neq r$ to the $r^{th}$ network. The term $\beta_r$ represents a vector of input coefficients for the $r^{th}$ network. The existing literature assumes that $\lambda_{rk} = 0$ for all $k \neq r$.

We allow the within-network and cross-network effects to be positive or negative, but for convenience we will refer to $W_{rr}$ as the peer network and $W_{rk}$ as the competitor network.

There are other interpretations of the model. For example, we can think of the index $r$ as representing distinct markets, where the networks might compete. If $\lambda_{rk} \neq 0$ for all $r, k$, then all networks compete in all markets. If any network $r$ does not compete in market $k \neq r$, then $\lambda_{rk} = 0$. In our application to the NBA, networks are individual teams, and in any game there are only ever two networks at a time on the right-hand side of the model: the

---

2When $\lambda_{rk} = 0$ for all $k \neq r$, the model reduces to the Horrace, Liu, and Patachini (2016) model.
peer network and the opposing teams network. We can also think of the model being for a single firm with different simultaneous projects \( r = 1, \ldots, R \). In this case, there might only be one manager making decisions for each project (network), but that presents no additional difficulties in what follows.

Then, the outcome function of the system of \( R \) equations in a matrix form is

\[
Y = \sum_{r=1}^{R} \sum_{k=1}^{R} \lambda_{rk} G_{rk} Y + XB + U
\]

where \( Y = (y'_1, \ldots, y'_R)' \), \( X = \text{Diag}(x'_1, \ldots, x'_R)' \), \( U = (u'_1, \ldots, u'_R)' \) and \( B = (\beta'_1, \ldots, \beta'_R)' \). \( G_{rk} \) is an \( N \times N \) block matrix with \( R \) row blocks and \( R \) column blocks. The blocks in \( G_{rk} \) are all blocks of zeros except for \( r^{th} \) row block in the \( k^{th} \) column block position, which equals \( W_{rk} \). For example, if there are \( R = 2 \) networks, then \( G_{11} = [\begin{bmatrix} W_{11} & 0 \\ 0 & 0 \end{bmatrix}] \), \( G_{22} = [\begin{bmatrix} 0 & 0 \\ 0 & W_{22} \end{bmatrix}] \), \( G_{12} = [\begin{bmatrix} 0 & W_{12} \\ 0 & 0 \end{bmatrix}] \), and \( G_{21} = [\begin{bmatrix} 0 & 0 \\ W_{21} & 0 \end{bmatrix}] \) where 0 is a conformable matrix of zeros. If all the networks are the same size \( n_r = n \), then \( G_{rk} = g_{rk} \otimes W_{rk} \), where \( g_{rk} \) is an \( R \times R \) matrix of zeros except for a ‘1’ in the \( r^{th} \) row and \( k^{th} \) column. In our NBA application, all the networks are the same size (5 x 5) with each row representing an active player. This representation is similar to the higher-order SAR model of Lee and Liu (2010) but without autoregressive errors. We assume that \( I - \sum_{r=1}^{R} \sum_{k=1}^{R} \lambda_{rk} G_{rk} \) is invertible so that the model is in equilibrium. Then, the reduced form of equation (2.2) is

\[
Y = (I - \sum_{r=1}^{R} \sum_{k=1}^{R} \lambda_{rk} G_{rk})^{-1}(XB + U).
\]

The model forms a simultaneous SAR equation system (or network model with heterogeneous interaction effects) where the simultaneities are coming not only from the peer network but also from the competitor networks. Therefore, this model can be viewed as an extension of a single equation SAR model to a simultaneous system of multiple SAR equation models. Similar simultaneous models exist in the literature, but these models are meant for simultaneous outcome variables, rather than for a single outcome variable with simultaneous networks in a competitive setting. Kelejian and Prucha (2004) generalize the single equation SAR
model due to Cliff and Ord (1973, 1981) to a system of spatially interrelated cross sectional equations, and they consider two stage least square (2SLS) and three stage least square (3SLS) estimation methods. Baltagi and Deng (2015) extend their model to a panel setting with random effects, while Cohen et al. (2017) extend it to a simultaneous equation network model with network fixed effects. They both consider 3SLS to estimate their models. Yang and Lee (2017) study identification and QML estimation for the model of Kelejian and Prucha (2004). All of these models focus on the cases where the cross interaction effects work through a single network or non-network structure, while our model contains different network matrices, allowing for a rich set of heterogeneous network interactions.

For empirical studies using the simultaneous SAR model, see Jeanty et al. (2010), Gebremariam et al. (2011), Baltagi and Bresson (2011), or Allers and Elhorst (2011).

Horrace et al. (2016) considers a single-equation version of this model where \( R = 1 \), implying no cross-network interactions. To understand the effect of ignoring the cross-networks, suppose that the correct model has \( R = 2 \), but we estimate the \( R = 1 \) model. For simplicity, assume that \( \lambda_{11} = \lambda_{22} = \lambda_1, \lambda_{12} = \lambda_{21} = \lambda_2, \beta_1 = \beta_2 = \beta \), and let \( G_1^* = G_{11} + G_{22} \), and \( G_2^* = G_{12} + G_{21} \). Suppose we assume that \( E(U|\{G_{rk}\}, X) = 0 \). Consider estimating the model omitting the cross network effects using 2SLS with a set of instrumental variables \( Z = [G_1^*X, X] \) without \( G_2^*X \). Then, \( \hat{\theta} = (\hat{\lambda}_1, \hat{\beta})' = (M'P_ZM)^{-1}M'P_Z Y \) where \( M = [G_1^*Y, X] \) and \( P_Z = Z(Z'Z)^{-1}Z' \). As the true model is \( Y = M\theta + \lambda_2G_2^*Y + U \), we have \( \hat{\theta} = \theta + (M'P_ZM)^{-1}M'P_Z(\lambda_2G_2^*Y + U) \). So \( \hat{\theta} - \theta = \lim_{N \to \infty}(\bar{M}'P_Z\bar{M})^{-1}\bar{M}'P_Z \lambda_2G_2^*\left(I - \sum_{i=1}^{2}\lambda_iG_i^*\right)^{-1}X\beta \) where \( \bar{M} = [G_1\left(I - \sum_{i=1}^{2}\lambda_iG_i\right)^{-1}X, \beta, X] \). Therefore, as long as \( \beta \neq 0 \) and \( \lambda_2 \neq 0 \), the \( \lambda_1 \) and \( \beta \) will be estimated inconsistently.

\(^{3}\)Kelejian and Prucha (2004) include a footnote that their model can be generalized to the case with simultaneous weight matrices which are unique to each variable, however, this is still more restrictive than our network interaction specification. Baltagi and Deng (2015) use two different weighting matrices for each variable, but they don’t have the cross-network effects.
Bias due to strategic interactions: Our model has endogenous variables, $Y$ on the right hand-side, which will be handled by standard spatial econometrics techniques. In addition to this source of endogeneity, the managers’ strategic actions may be correlated with the outcome. If this is the case, $E(U|\{G_r\}, X) \neq 0$ in equation (2.2). To address this issue, we set up a static game to formulate and correct the bias.\(^4\) We follow and adapt the basic methodologies in the game theory literature for static games of incomplete information with multiple equilibria. In particular, our arguments follow Bajari et al. (2010).\(^5\)

Each network $r$ has a network manager who takes actions from his finite and discrete choice set of strategies, $A = (1, \ldots, s, \ldots)$, on behalf of their members, and their decisions are based on two types of state variables, $(Z, e_r)$, where $Z$ is a vector of observable state variables (i.e., market characteristics) which are common to all the networks, and $e_r = (e_r(1), \ldots, e_r(s), \ldots)$ is $r$’s unobservable action-specific state variable.\(^6\) The game proceeds as follows: First, the state variables, $(Z, \{e_r\}_{r=1}^R)$, are realized. Then, the network managers simultaneously choose their actions from their choice set. Under the chosen action, networks produce single outcome, $Y$.

Let the $r^{th}$ network manager’s additively separable payoff function be $\pi_r(s_r, s_{-r}, Z, e_r(s)) = \pi_0^r(s_r, s_{-r}, Z) + e_r(s)$, where $s_{-r} = (s_1, \ldots, s_{r-1}, s_{r+1}, \ldots, s_R)$, the collection of all the network managers’ decisions except $r$. We assume error terms $\{e_r(s)\}_{r=1}^R$ are distributed iid across $s$ and network $r$ with distributions $\{G_{e_r}(\cdot)\}_{r=1}^R$. We assume that $Z, \{\pi_0^r(\cdot)\}_{r=1}^R$ and $\{G_{e_r}(\cdot)\}_{r=1}^R$ are common knowledge to the managers while the $\{e_r(s)\}$ are private information. Then, we can define $r$’s information set as $\eta_r = \{Z, e_r, \{\pi_0^r(\cdot)\}_{r=1}^R, \{G_{e_r}(\cdot)\}_{r=1}^R\}$ and $r$’s decision as $\rho_r(\cdot) : \eta_r \rightarrow A$. Under these settings, the conditional choice probability (CCP) of $r$ choosing

\(^4\)There may be other econometric remedies to address the endogeneity issue. Recently, two categories of methodologies have been proposed to address the endogeneity in formation of spatial or network links: One is (Bayesian) One step Full information approach by Goldsmith and Imbens (2013) and Hsieh and Lee (2016), and the other is Multiple step Control Function approach by Qu and Lee (2015) and Horrace et al. (2016).

\(^5\)Generalizing the following model to a dynamic game is left for future research.

\(^6\)Time period, $t$, is suppressed here. Follow Bajari et al. (2010), without loss of generality we assume that the strategies of the network managers are the same.
s ∈ A at a given realization of the Z is given by

\[ \delta_r(s|Z) = \int 1\{\rho_r(\eta_r) = s\} dG_{e_r}(e_r) \]  

(2.3)

which can be interpreted as the beliefs formed by r’s opponents regarding r’s decision. Since network manager r does not know the other managers’ decisions at the time of her decision, her strategy is based on her expected payoff

\[ \pi_r^{ex}(s, Z, e_r(s)) = \sum_{s_{-r}} \prod_{k \neq r} \delta_k(s_k|Z) \pi_r(s, s_{-r}, Z) + e_r(s) = \varphi(s, Z) + e_r(s) \]  

(2.4)

We can see that the expected payoff function is similar to the standard random utility model. The only difference is that the probability distributions over other managers’ actions are affecting manager r’s utility. Then, it is straightforward that s will be chosen for a given realization of Z if only if

\[ \pi_r^{ex}(s, Z, e_r(s)) > \max_{s' \neq s} \pi_r^{ex}(s', Z, e_r(s')) \Leftrightarrow \varphi(s, Z) + e_r(s) > \max_{s' \neq s} \varphi(s', Z) + e_r(s') \]  

(2.5)

for s, s' ∈ A. It follows immediately that \( \delta_r(s|Z) = Pr(\max_{s' \neq s} \varphi(s', Z) - \varphi(s, Z) + e_r(s') - e_r(s) < 0) \) in equilibrium (e.g. Bayesian-Nash Equilibrium).^7

**Formulation of the selection bias:** We formulate the bias by assuming the error terms, \((u_r, e_r)\), from the outcome equation and the payoff function (respectively) are statistically dependent, a standard approach. If there’s a correlation between the two errors, the expectation of \( u_r \) conditional on the choice of strategy, s will not have a zero mean; that is,

\[ E(u_r|s) = E(u_i|\max_{s' \neq s} \varphi(s', Z) - \varphi(s, Z) + e_r(s') - e_r(s) < 0) \neq 0 \]  

(2.6)

---

^7 In the game literature the focus is often to estimate the payoff function, and doing so requires additional structure be imposed on the function. However, this is not the focus here, so additional structure is not necessary.
This correlation may exist when the two errors contain a common component, unobserved by the econometrician (e.g., a network-strategy specific fixed effect in the outcome equation and the strategy selection equation).

The bias in the outcome equation can be expressed parametrically or semi-parametrically following Lee (1983) or Dahl (2002).

1. **Lee’s approach**: From (2.5), we see that strategy \( s \) will be chosen by manager \( r \) if only if \( \epsilon^*_r < 0 \), where \( \epsilon^*_r = \max_{s' \neq s} \varphi(s', Z) - \varphi(s, Z) + e_r(s') - e_r(s) \). \( \epsilon^*_r \) is a new random variable with a distribution of \( F_r \). Following Lee (1983) and Horrace, Liu, and Patacchini (2016), we can reduce the dimensionality of \( F_r \) by the transformation \( J_r(\cdot) \equiv \Phi^{-1}(F_r(\cdot)) \), where \( \Phi^{-1} \) is the inverse of the standard normal CDF. Then, \( J_r(\epsilon^*_r) \) becomes a standard normal random variable. For notational simplicity, let \( J_r(\epsilon^*_r) \equiv \xi_r \).

We further assume \( \epsilon_r \) and \( u_{r,i} \) for \( i = 1, \ldots, n_r \) are i.i.d with a joint normal distribution,

\[
\begin{bmatrix}
  u_{r,i} \\
  \epsilon_i
\end{bmatrix} = N \left( 
  \begin{bmatrix}
    0 \\
    0
  \end{bmatrix}, 
  \begin{bmatrix}
    \sigma^2_u & \sigma_{12} \\
    \sigma_{12} & 1
  \end{bmatrix}
\right),
\]

where \( u_{r,i} \) is the \( i^{th} \) element of the vector \( u_r \). Then it can be shown that

\[
E(u_r|s) = E(u_r| \epsilon_r < J_r(0), Z) = -\sigma_{12} \frac{\phi(\Phi^{-1}\delta_r(s|Z))}{\delta_r(s|Z)} \iota_{n_r} \tag{2.7}
\]

where \( \delta_r(s_r|Z) \) is the selection probability from (2.3) and \( \iota_{n_r} \) is an \( n_r \)-dimensional vector of 1’s.

2. **Dahl’s approach**: Without a parametric distributional assumption on \( u_r \) and \( \epsilon_r \), we make the index sufficiency assumption of Dahl (2002) such that

\[
h(\epsilon_r, u_r|\{\varphi(s, Z)\}_{s \in A}) = h(\epsilon_r, u_r|\delta_r(s|Z)) \tag{2.8}
\]

where \( h(\cdot, \cdot) \) is some bivariate distribution for \( u_r \) and \( \epsilon_r \). As Dahl points out, this
assumes that the selection probability $\delta_r(s|Z)$ exhausts all the information about the behaviors of the two errors. Then, the bias will be given by $E(u_r|s) = \psi_r(\delta_r(s|Z))$, where $\psi_r(\cdot)$ is an unknown function that can be estimated nonparametrically.

With these approaches, we can identify the bias in the outcome equation by rewriting (2.1) as

$$y_r = \lambda_{rr} W_{rr} y_r + \sum_{k \neq r} \lambda_{r,k} W_{r,k} y_k + x_r \beta_r + a_{r,t_n} + u_r^*$$  \hspace{1cm} (2.9)

where $a_{r,t_n} = E(u_r|s)$ are network specific fixed-effects due to the strategic actions of the network managers, and $u_r^* = u_r - E(u_r|s)$ with zero mean by construction, so the endogeneity disappears in the outcome equations conditional on $a_r$.

2.2.2 Estimation

We have shown that the strategic bias can be reduced to a network specific fixed effect. If our primary interest is just to estimate network effects, it is sufficient to use the within transformation around each network to remove the network specific fixed effects, and, then, apply QML or GMM to estimate the model. We will illustrate this case in this section. However, if there are network invariant regressors (e.g., network specific characteristic) and we are also interested in estimating the strategic bias, we have to consider one more step as the within transformation eliminates them in the first step. We will briefly discuss the estimation procedure for this case at the end of this section.

Here, we focus on QML to estimate the model and provide a sketch of the GMM method using the $R = 2$ networks case in the Appendix B. After accounting for the strategic bias,

---

8The conceptual foundations for correcting endogeneity in this way for a social interactions model can be traced to a series of papers by Brock and Durlauf (2001, 2002, 2006).

9In our NBA application, we only estimate network peer-effects and marginal effects for network varying regressors.

10One may also consider GS2SLS or GS3SLS to estimate the models (Kelejian and Prucha, 2004; Lee, 2003).
the complete system is

\[ Y = \sum_{r=1}^{R} \sum_{k=1}^{R} \lambda_{rk} G_{rk} Y + XB + A + U^* \]  

(2.10)

where \( A = (a_1^\prime, \ldots, a_R^\prime) \) and \( U^* = (u_1^*, \ldots, u_R^*) \). To remove the bias term \( A \) and avoid the incidental parameter problem (Neyman and Scott, 1948), we transform with projector \( J_Q = \text{Diag}(Q_1, \ldots, Q_R) \) where \( Q_r \) for \( r = 1, \ldots, R \) is the within transformation matrix, \( Q_r = I_n - \frac{1}{n} t_{n_r} t_{n_r}^\prime \). Then, as \( Q_r t_{n_r} = 0 \) and \( Q_r u_r^* = Q_r u_r \), we have

\[ J_Q Y = J_Q \sum_{r=1}^{R} \sum_{k=1}^{R} \lambda_{rk} G_{rk} Y + J_Q XB + J_Q U \]  

(2.11)

Extensive study of identification conditions for network models and multivariate SAR models can be found in Bramoulle et al (2009), Cohen et al (2017) and Yang and Lee (2017). In Appendix B, we discuss identification conditions for (2.11). The identification condition will be generally satisfied because we have multiple sets of network matrices and exogenous regressors from each group, which produces enough variation to identify the coefficients in our model.

We assume that each element \( u_{r,i} \) for \( i = 1, \ldots, n_r \) is \( \text{iid}(0, \sigma^2_r) \). We note that the disturbances, \( Q_r u_r \), in (2.11) are linearly dependent because the variance matrix \( \sigma^2_r Q_r \) is singular. Following Lee et al. (2010a), we consider an equivalent but more effective transformation which can eliminate the network fixed-effects while maintaining interdependency between the errors.
disturbances. Let the orthonormal matrix of $Q_r$ be $[P_r, t_{n_r}/\sqrt{n}]$. The columns in $P_r$ are
eigenvectors of $Q_r$ corresponding to the eigenvalue one, such that $P'_r t_{n_r} = 0$, $P'_r P_r = I_{n_r-1}$ and$P_r P'_r = Q_r$. Then, premultiplying (2.10) by $J_p = Diag (P'_1, \ldots, P'_R)$ leads to

\[
J_p Y = J_p \sum_{r=1}^R \sum_{k=1}^R \lambda_{rk} G_{rk} Y + J_p X B + J_p U \tag{2.12}
\]

Let $\bar{Y} = J_p Y_m$, $\bar{X} = J_p X$, $\bar{U} = J_p U$, $\bar{G}_{rk} = J_p G_{rk} J_p$. Due to $J_p G_{rk} = \bar{G}_{rk} J_p$,\footnote{\bar{G}_{rk} J_p = J_p G_{rk} J_Q = J_p G_{rk} (I - \text{diag}(t_{n_1} t_{n_1}' / n_1, \ldots, t_{n_R} t_{n_R}' / n_R)) = J_p G_{rk}$ because $G_{rk}$ is row-normalized so $J_p G_{rk} \text{diag}(t_{n_1} t_{n_1}' / n_1, \ldots, t_{n_R} t_{n_R}' / n_R) = 0$.} then this implies

\[
\bar{Y} = \sum_{r=1}^R \sum_{k=1}^R \lambda_{rk} \bar{G}_{rk} \bar{Y} + \bar{X} B + \bar{U} \tag{2.13}
\]

Note that $\bar{u}_r$ for $r = 1, \ldots, R$ is now $u_r \sim (0, \sigma^2_r I_{n_r-1})$. Therefore, the likelihood function\footnote{The likelihood is conditional likelihood because it is conditional on the sufficient statistic, the mean of $y_r$. Lee (2007a).} is

\[
\ln L(\Lambda, B, \Sigma) = -\sum_{r=1}^R \frac{n_r - 1}{2} \ln (2\pi \sigma^2_r) + \ln |\bar{S}(\Lambda)| - \sum_{r=1}^R \frac{\bar{e}_r(\theta_r)' \bar{e}_r(\theta_r)}{2\sigma^2_r} \tag{2.14}
\]

where $\Lambda = (\Lambda'_1, \ldots, \Lambda'_R)$ with $\Lambda_r = (\lambda_{r,1}, \ldots, \lambda_{r,R})$, $\Sigma = (\sigma^2_1, \ldots, \sigma^2_R)$, $\bar{S}(\Lambda) = l - \sum_{r=1}^R \sum_{k=1}^R \lambda_{rk} \bar{G}_{rk}$, and $\bar{e}_r(\theta_r) = \bar{y}_r - \sum_{k=1}^R \lambda_{rk} \bar{W}_{rk} \bar{y}_k - x_r \beta_r$ where $\theta_r = (\Lambda_r, \beta_r)$, and $\bar{W}_{rk}$ and $\bar{x}_r$ are defined similarly as in (2.13). From Lemma \ref{lemma:B.3.1} in the Appendix B, we show $\ln |\bar{S}(\Lambda)| = -\ln f(\Lambda) + \ln |S(\Lambda)|$ where $S(\Lambda) = l - \sum_{r=1}^R \sum_{k=1}^R \lambda_{rk} G_{rk}$ and $f(\Lambda)$ is some function of $\Lambda$. For example, when $R = 2$, $f(\Lambda) = (1 - \lambda_{11})(1 - \lambda_{22}) - \lambda_{12} \lambda_{21}$ and when $R = 3$, $f(\Lambda) = (1 - \lambda_{11})(1 - \lambda_{22})(1 - \lambda_{33}) - (1 - \lambda_{11}) \lambda_{23} \lambda_{32} - (1 - \lambda_{22}) \lambda_{13} \lambda_{31} - (1 - \lambda_{33}) \lambda_{21} \lambda_{12} - \lambda_{13} \lambda_{21} \lambda_{32}$. Using this result, we can evaluate the likelihood without $P_r$ as

\[
\ln L(\Lambda, B, \Sigma) = -\sum_{r=1}^R \frac{n_r - 1}{2} \ln (2\pi \sigma^2_r) - \ln f(\Lambda) + \ln |S(\Lambda)| - \sum_{r=1}^R \frac{\epsilon_r(\theta_r)' Q_r \epsilon_r(\theta_r)}{2\sigma^2_r} \tag{2.15}
\]

where $\epsilon_r(\theta_r) = y_r - \sum_{k=1}^R \lambda_{rk} W_{rk} y_k - x_r \beta_r$. There are two things to note here. First, we need to restrict the parameter space for $\Lambda$ to guarantee that $|S(\Lambda)|$ and $f(\Lambda)$ are strictly positive so
that the likelihood is well defined. From Lee and Liu (2010), the parameter space for $|S(\Lambda)|$ is strictly positive, when $\sum_{r=1}^{R} \sum_{k=1}^{R} |\lambda_{rk}| < 1$, as our network matrices are row-normalized. However, Elhorst, Lacombe and Piras (2012) argue that this may be too restrictive, and suggest a stationary-region search methodology for $\Lambda$ that is potentially less so, while ensuring a well-defined likelihood function. Second, it may be difficult to evaluate $|S(\Lambda)|$. The Ord (1975) eigenvalue device may used to compute the determinant; however, it may only work when the number of networks is small and all the network matrices are sparse. If the number of networks is large, then GMM may be preferred to QML, as it avoids the computational difficulties of evaluating the determinant.

To simplify estimation, we concentrate out $B$ and $\Sigma$ in (17). The QMLE of $\beta_r$ and $\sigma_r^2$, given $\Lambda_r$ is $\hat{\beta}_r(\Lambda_r) = (x'_rQ_rx_r)^{-1}x'_rQ_r\mu_r(\Lambda_r)$ where $\mu_r(\Lambda_r) = y_r - \sum_{k=1}^{R} \lambda_{rk}W_{rk}y_k$, and

$$\hat{\sigma}_r^2(\Lambda_r) = \frac{\epsilon'_r(\theta_r)Q_r\epsilon_r(\theta_r)}{n_r - 1} = \mu_r(\Lambda_r)'[Q_r - Q_r x_r (x'_rQ_rx_r)^{-1}x'_rQ_r]\mu_r(\Lambda_r).$$  \hspace{1cm} (2.16)

Then the concentrated log-likelihood function in $\Lambda$ is

$$\ln L^c(\Lambda) = -\sum_{r=1}^{R} \frac{n_r - 1}{2} [\ln(2\pi) + 1] - \ln f(\Lambda) + \ln |S(\Lambda)| - \sum_{r=1}^{R} \frac{n_r - 1}{2} \ln \hat{\sigma}_r^2(\Lambda_r).$$  \hspace{1cm} (2.17)

Then the QMLE, $\hat{\Lambda}$, is the maximizer of the likelihood, and the QMLE of $B$ and $\Sigma$ are $\hat{\beta}_r(\hat{\Lambda}_r)$ and $\hat{\sigma}_r^2(\hat{\Lambda}_r)$ for $r = 1, ..., R$, respectively. The asymptotic distribution for these estimators can be derived from Lee et al. (2010a, Appendix B) with appropriate modifications.

**Estimation of the strategic bias**

The most disaggregate level of variability in this model is the individual worker-level, $i = 1, ..., n_r$, and this is the variability that ultimately identifies the peer and competitor effects, $\lambda_{rk}$. However, there may be columns in $x_r$ that vary at the network level (or higher), and these columns will necessarily be eliminated by the within-network transformation of the
data. If we are only interested in estimating the peer- and competitor-effects, this will be fine.\footnote{This is our approach in the empirical exercise in the next section.} However, if we are interested in estimating the strategic bias caused by the network managers (and the coefficients on network-varying exogenous variables), we need to consider one more step as follows:

1. Estimate the selection probability $\tilde{\delta}_i(s|Z)$ for $i = r, ..., R$ using a nonparametric or parametric model (kernel smoothing, local polynomial regression, or the logit model).\footnote{Alternatively, we can simply use the cell statistics discussed in Dahl (2002).}

2. Using $\hat{\Lambda}$ and $\hat{B}$ from (2.15), compute the residual $\hat{\upsilon}_r = \iota_n'(y_r - \sum_{k=1}^R \hat{\lambda}_{rk} W_{rk} y_k - x_r \hat{\beta}_r)/n_r$ for $r = 1, ..., R$. Let $\beta_{r,2} \subseteq \beta_r$ be the coefficients on network-level varying regressors, $x_{r,2} \subseteq x_r$. Then the bias and $\beta_{r,2}$ can be estimated from the OLS regression $\hat{\upsilon}_r = x_{r,2} \beta_{r,2} + \mu(\hat{\delta}_i(s|Z)) + \xi_r$ where $\xi_r$ is an i.i.d. error term and $\mu(\hat{\delta}_i(s|Z))$ is either given by $-\sigma_{12} \frac{\phi^{-1}(\hat{\hat{\delta}_i(s|Z)})}{\hat{\delta}_i(s|Z)}$ (Lee’s approach) or some nonparametric, single-index formulation (based on Dahl’s approach). See Horrace, Liu, and Patacchini (2016) for an explanation of Dahl’s approach in this context.

### 2.3 Empirical application

#### 2.3.1 Empirical model and variables

In this section, we apply our network competition model to NBA data for 30 teams over the 2013-14, 2014-15, and 2015-2016 regular seasons. The primitive play-by-play data were purchased and downloaded from BigDataBall.com. We then formatted the data to the player-period level, where a period represents any contiguous game period in which the same ten players are on the court. This formatting is similar to that done in the calculation of the player statistic real plus minus. We tabulate player box-score data to obtain wins produced (see, e.g., Berri, 1999) at the time-period level. The league plays 1,230 regular-season games...
per season (41 home games for each of 30 teams per regular season). Therefore, our data spans 3,690 regular-season games, consisting of roughly 30 time periods per game. This produces 112,204 time periods in which we observe the play of 10 players at a time, producing a total of 1,122,040 observations. In each game a coach typically has 15 players to fill a network of five players at a time. Following Horrace, Liu, and Patachini (2016), we drop time periods less than 30 seconds and overtime periods. This results in 83,334 time periods for the league, which is equivalent to an average of about 926 periods per team-season.

Our outcome variable is *wins produced*, a continuous weighted average of individual player offensive and defensive statistics that will be defined in what follows. *Wins produced* is highly predictive of team success and is separable (i.e., measurable at the individual level). The variable is a version of the outcome variable in Horrace, Liu, and Patachini (2016) but with different weights on the component statistics in the weighted average, as we shall see. We calculate this outcome variable for each player in each period of the data. Overall, our data and specification are designed to mimic the Horrace, Liu, and Patachini (2016) analysis of the Syracuse Men’s Basketball team.

**PRODUCTION FUNCTION:** With thirty teams in the league, there is scope for estimating $30^2$ peer- and competitor-effects in our model. Unfortunately, the number of games between a given pair of teams is small (three or four at maximum), so there is not enough head-to-head data to estimate this many parameters. Consequently, we assume a given team’s cross-network competitor-effects from the 29 other teams are the same. That is, for team $r$, $\lambda_{rk} = \lambda_r$ for $k \neq r$. Conceptually, our restriction on the cross-network competitor-effects is equivalent to team $r$ playing a season long game against all the other teams in the league, consisting of subgames against individual teams (and coaches), and where at the end of each subgame, the outcome variable, $y_{rt}$ is set to zero for each player. In other words, team $r$ has a representative

---

16Understanding the effect of player injuries (or player ineligibility) on the coaches’ decisions is left for future research. Sports injuries are analogous to worker absenteeism.
opponent, $k$, and opposing teams take turns being that representative opponent. Consequently, interpretation of the competitor-effect will be challenging, so we focus the analysis on the within-network effects. The unique structure of competition in this setting, requires slightly different notation than the more general model. The production function for team $r$ and $k$ in period $t$ of game $s$ is

$$
\begin{align*}
y_{rts} &= \lambda_r W_{rts} y_{rts} + \lambda_r W_{knts} y_{kts} + x_{rts} \beta_r + u_{rts} \\
y_{kts} &= \lambda_k W_{knts} y_{kts} + \lambda_k W_{knts} y_{rts} + x_{kts} \beta_k + u_{kts}
\end{align*}
$$

(2.18)

where $y_{rts}$ and $y_{kts}$ are the $5 \times 1$ outcome vector of team $r$'s and team $k$'s chosen lineup in period $t$ of game $s$, respectively, and $W_{rts}$ and $W_{knts}$ are the $5 \times 5$ zero diagonal and row-normalized matrices for the within-network interactions, and $W_{knts}$ and $W_{knts}$ are similarly defined matrices for cross-network interactions. The $x_{rts}$ and $x_{kts}$ are matrices of the player-varying exogenous variables for team $r$ and $k$'s lineup in period $t$ of game $s$, respectively.

The $u_{rts}$ and $u_{kts}$ are $5 \times 1$ error term vectors, in which each element is assumed to be $iid \ N(0, \sigma^2_r)$ and $iid \ N(0, \sigma^2_k)$, respectively. Following the methodology in section 2.2.2, the log-likelihood for the system of equations (2.18), denoted as $\ln L_{ts}$, is,

$$
\ln L_{ts} = - \sum_{h=r,k} 2 \ln(2\pi \sigma^2_h) - \ln f_s + \ln |S_{ts}| - \sum_{h=r,k} \frac{\epsilon'_{hts}(\theta_h)Q\epsilon_{hts}(\theta_h)}{2\sigma^2_h}
$$

(2.19)

where $f_s = [(1 - \lambda_{rr})(1 - \lambda_{kk}) - \lambda_r \lambda_k]$, and $S_{ts} = I_{10} - \lambda_{rr} G_{rts} - \lambda_r G_{knts} - \lambda_k G_{kts} - \lambda_{kk} G_{knts}$ with $G_{rts} = \begin{bmatrix} W_{rts} & 0 \\ 0 & 0 \end{bmatrix}$, $G_{knts} = \begin{bmatrix} 0 & W_{knts} \\ 0 & 0 \end{bmatrix}$, $G_{kts} = \begin{bmatrix} 0 & 0 \\ W_{knts} & 0 \end{bmatrix}$, $G_{kkts} = \begin{bmatrix} 0 & 0 \\ 0 & W_{knts} \end{bmatrix}$, $Q = I_5 - \frac{1}{2} I_5' I_5$, $\epsilon_{hts}(\theta_r) = y_{rts} - \lambda_r W_{rts} y_{rts} - \lambda_r W_{knts} y_{kts} - x_{rts} \beta_r$ with $\theta_r = (\Lambda_r, \beta_r)$ and $\epsilon_{hts}(\theta_k)$ is given similarly. The log-likelihood for entire season is just the sum of the likelihood in (2.19) over all games, $s$, and time periods, $t$.

\footnote{As previously noted, the $x_{rts}$ may contain columns that vary at the network-level (or higher), but they will be eliminated from the model with the within transformation. Their coefficients may be recovered in the estimation of the coach’s strategic bias, but we ignore them in our analysis because we are only concerned with estimating the peer- and competitor-effects, which are preserved under the within transformation.}
VARIABLES: We use the wins produced measure based on the work of sports economist David Berri (Berri, 1999; Berri et al. 2006): \( y_{irts} = (0.064 \times 3PT_{irts} + 0.032 \times 2PT_{irts} + 0.017 \times FT_{irts} + 0.034 \times REB_{irts} + 0.033 \times STL_{irts} + 0.020 \times BLK_{irts} - 0.034 \times MFG_{irts} - 0.015 \times MFT_{irts} - 0.034 \times TO_{irts}) / \text{Mins}_{irts} \), where \( 3PT_{irts} \), \( 2PT_{irts} \), and \( FT_{irts} \) are 3-point field goals made, 2-point field goals made, and free throws made, respectively, \( REB_{irts} \) is rebounds, \( STL_{irts} \) is steals, \( BLK_{irts} \) is blocks, \( MFG_{irts} \) is missed field goals, \( MFT_{irts} \) is missed free throws, \( TO_{irts} \) is turnovers, and \( Mins_{irts} \) is minutes played by player \( i \) of team \( r \) in period \( t \) of game \( s \). Wins produced per minute (or wins per minute) estimates a player’s marginal win productivity based upon player-level variables related to team-winning. It represents a leading measure of NBA player production.\(^{19}\)

\[^{18}\text{Horrace, Liu, and Patacchini (2016) use both same- and cross-type weight matrices. The different-type weight matrix is } W^d, \text{ where } W^d_{0,i} = [w_{0,ij}] \text{ is an adjacency matrix with } w_{0,ij} = 1 \text{ if the } i^{th} \text{ and } j^{th} \text{ players are a guard and a forward (or vice versa). Then row-normalize } W^d_0 \text{ so that } W^d1_N = 1_N. \text{ However, they noticed that including both weight matrices into the model at the same time may lead to a multicollinearity problem. Therefore, we only include the same type weight matrix in this exercise.}\]

\[^{19}\text{See www.basketball-reference.com/about/bpm.html or wagesofwins.com/how-to-calculate-wins-produced/}\]
Experience_{irts} and Fatigue_{irts}. Experience_{irts} is minutes played from the start of the game to the end of period t-1, and Fatigue_{irts} is minutes continuously played until the end of period t-1. We also included player dummies to control for player-specific heterogeneity. The descriptive statistics for the continuous variables are presented in the tables of Appendix B.

2.3.2 Estimation strategy

We estimated the model for each season after concentrating out $B$ and $\Sigma$. However, optimizing the entire log-likelihood with respect to 60 parameters for 30 teams at the same time may not be efficient. So, we employed the estimation strategy below:

1. **Preliminary estimation**: We include the regressor for the cross team interaction, such as $W_{kt}y_{kts}$ or $W_{kt}y_{rts}$ in (2.18), but do not account for the endogeneity from the regressors. That is, we assume $\lambda_r = \lambda_k = 0$ in $\ln[(1 - \lambda_r)(1 - \lambda_k) - \lambda_r\lambda_k]$ and $\ln |I_{10} - \lambda_r G_{rts} - \lambda_r G_{kt} - \lambda_k G_{kts} - \lambda_k G_{kts}|$ of (2.19). Then, the entire log-likelihood for 30 teams can be separated into each individual team’s log-likelihood. Then, we can optimize each individual team’s log-likelihood separately to get initial estimates for $\lambda_{hh}$ and $\lambda_h$ for $h = 1, ..., 30$. We denote the initial estimates as $\hat{\lambda}_{hh}^{pre}$ and $\hat{\lambda}_h^{pre}$ for $h = 1, ..., 30$.

2. **Updating**: We update $\hat{\lambda}_{hh}^{pre}$ and $\hat{\lambda}_h^{pre}$ for $h = 1, ..., 30$ using the original entire log-likelihood, but this update is carried out team-by-team. That is, we optimize the entire log-likelihood w.r.t one of the teams’ $\Lambda$ given all other teams’ $\Lambda$ from the preliminary estimation. We continue these updates from the first team to the last team (the order for discussions of wins produced. The NBA scales this statistic to the game level by multiplying by 48 minutes per game. It is typically reported at the player level but we report it the team level in Appendix B.

20If $\lambda_r$ and $\lambda_k$ are zero, it is obvious that $\ln [(1 - \lambda_r)(1 - \lambda_k) - \lambda_r\lambda_k] = \ln(1 - \lambda_r) + \ln(1 - \lambda_k)$. This is also true for $\ln |I_{10} - \lambda_r G_{rts} - \lambda_r G_{kt} - \lambda_k G_{kts} - \lambda_k G_{kts}|$ because it becomes a block diagonal matrix when $\lambda_r$ and $\lambda_k$ are zero.
3. **Convergence**: We iterate the set of updates for 30 teams until the difference in $\hat{\Lambda}$ between two consecutive iterations is below some threshold\(^{21}\)

### 2.3.3 (Preliminary) Estimation results

We now present preliminary results, focusing only on the peer-effect estimates. (Estimation results for the complete system will be added in a subsequent revision.) Estimation results for the outcome equation can be found in Appendix B. We summarize the main results on team-level peer-effects in Table 2.1.

Peer-effects measure team chemistry conditional on strategies, abilities and opposition and do not measure team quality. Like a talented shooter can play well even with sub-optimal shot selection, a talented team can perform well even given low peer-effects. Table 2.1 contains the ranked peer-effects for 30 NBA teams in each of three seasons. Bounded on the unit-circle, a peer-effect close to 1 ($-1$) indicates good (poor) conditional team chemistry, as player performance is positively (negatively) linked to average teammate performance. Consider the results in Table 2.1 on the 2013-14 season, where Utah (UTA) had the largest peer-effect of 0.045. That is, when the team’s average “wins produced” increases, the team gains an additional 4.5% by virtue of its good chemistry, conditional on coaching strategy and other environmental and performance variables. In this sense the peer-effect is like an output multiplier, and teams with large positive values benefit from their own team chemistry. In the 2013-14 season Utah had an average wins produced per minute of 0.0049 with a standard deviation of 0.0248. (See Appendix B, Table B.1) The distribution of this is in Figure 2.1. This is the distribution over active Utah players for all time periods in the 2013-14 season. Utah’s 2013-14 “wins produced” is highly variable with a small mean, as the team a) played in 1,820

\(^{21}\)In our exercise, we use a criterion of $||\hat{\Lambda} - \hat{\Lambda}^0||_2 < 10^{-8}$ where $\hat{\Lambda}$ is an estimate for $\Lambda$ from the current iteration and $\hat{\Lambda}^0$ is the estimate from the previous iteration.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Peer-effect</th>
<th>Team</th>
<th>Peer-effect</th>
<th>Team</th>
<th>Peer-effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UTA</td>
<td>0.045***</td>
<td>IND</td>
<td>0.054***</td>
<td>CHI</td>
<td>0.065***</td>
</tr>
<tr>
<td>2</td>
<td>BOS</td>
<td>0.044***</td>
<td>PHX</td>
<td>0.033**</td>
<td>SAS</td>
<td>0.052***</td>
</tr>
<tr>
<td>3</td>
<td>PHX</td>
<td>0.041***</td>
<td>UTA</td>
<td>0.032***</td>
<td>MIN</td>
<td>0.049***</td>
</tr>
<tr>
<td>4</td>
<td>WAS</td>
<td>0.031**</td>
<td>PHI</td>
<td>0.023*</td>
<td>MIL</td>
<td>0.046***</td>
</tr>
<tr>
<td>5</td>
<td>IND</td>
<td>0.027**</td>
<td>MIN</td>
<td>0.020*</td>
<td>MEM</td>
<td>0.045***</td>
</tr>
<tr>
<td>6</td>
<td>MIN</td>
<td>0.026**</td>
<td>CLE</td>
<td>0.018</td>
<td>ORL</td>
<td>0.037***</td>
</tr>
<tr>
<td>7</td>
<td>DAL</td>
<td>0.022*</td>
<td>WAS</td>
<td>0.018</td>
<td>BKN</td>
<td>0.035***</td>
</tr>
<tr>
<td>8</td>
<td>OKC</td>
<td>0.021*</td>
<td>LAC</td>
<td>0.017</td>
<td>DET</td>
<td>0.030**</td>
</tr>
<tr>
<td>9</td>
<td>SAS</td>
<td>0.021*</td>
<td>ORL</td>
<td>0.014</td>
<td>SAC</td>
<td>0.028**</td>
</tr>
<tr>
<td>10</td>
<td>TOR</td>
<td>0.020*</td>
<td>SAC</td>
<td>0.011</td>
<td>POR</td>
<td>0.027**</td>
</tr>
<tr>
<td>11</td>
<td>BKN</td>
<td>0.019</td>
<td>SAS</td>
<td>0.009</td>
<td>BOS</td>
<td>0.022*</td>
</tr>
<tr>
<td>12</td>
<td>CLE</td>
<td>0.018</td>
<td>MEM</td>
<td>0.008</td>
<td>TOR</td>
<td>0.020*</td>
</tr>
<tr>
<td>13</td>
<td>PHI</td>
<td>0.017</td>
<td>BKN</td>
<td>0.007</td>
<td>UTA</td>
<td>0.016</td>
</tr>
<tr>
<td>14</td>
<td>NOP</td>
<td>0.017</td>
<td>NOP</td>
<td>0.006</td>
<td>CLE</td>
<td>0.015</td>
</tr>
<tr>
<td>15</td>
<td>DET</td>
<td>0.015</td>
<td>DEN</td>
<td>0.002</td>
<td>HOU</td>
<td>0.014</td>
</tr>
<tr>
<td>16</td>
<td>MIL</td>
<td>0.014</td>
<td>TOR</td>
<td>0.002</td>
<td>CHA</td>
<td>0.011</td>
</tr>
<tr>
<td>17</td>
<td>LAL</td>
<td>0.013</td>
<td>CHA</td>
<td>0.001</td>
<td>ATL</td>
<td>0.008</td>
</tr>
<tr>
<td>18</td>
<td>NYK</td>
<td>0.009</td>
<td>DAL</td>
<td>0.001</td>
<td>PHX</td>
<td>0.004</td>
</tr>
<tr>
<td>19</td>
<td>DEN</td>
<td>0.009</td>
<td>MIL</td>
<td>0.001</td>
<td>LAL</td>
<td>0.001</td>
</tr>
<tr>
<td>20</td>
<td>GSW</td>
<td>0.007</td>
<td>ATL</td>
<td>0.000</td>
<td>WAS</td>
<td>0.000</td>
</tr>
<tr>
<td>21</td>
<td>MEM</td>
<td>0.006</td>
<td>HOU</td>
<td>-0.001</td>
<td>NYK</td>
<td>0.000</td>
</tr>
<tr>
<td>22</td>
<td>CHA</td>
<td>0.004</td>
<td>LAL</td>
<td>-0.002</td>
<td>OKC</td>
<td>-0.002</td>
</tr>
<tr>
<td>23</td>
<td>SAC</td>
<td>0.000</td>
<td>POR</td>
<td>-0.003</td>
<td>LAC</td>
<td>-0.005</td>
</tr>
<tr>
<td>24</td>
<td>HOU</td>
<td>-0.002</td>
<td>CHI</td>
<td>-0.004</td>
<td>MIA</td>
<td>-0.007</td>
</tr>
<tr>
<td>25</td>
<td>ORL</td>
<td>-0.004</td>
<td>MIA</td>
<td>-0.005</td>
<td>NOP</td>
<td>-0.011</td>
</tr>
<tr>
<td>26</td>
<td>ATL</td>
<td>-0.005</td>
<td>DET</td>
<td>-0.006</td>
<td>PHI</td>
<td>-0.014</td>
</tr>
<tr>
<td>27</td>
<td>LAC</td>
<td>-0.009</td>
<td>OKC</td>
<td>-0.007</td>
<td>GSW</td>
<td>-0.014</td>
</tr>
<tr>
<td>28</td>
<td>POR</td>
<td>-0.010</td>
<td>GSW</td>
<td>-0.022**</td>
<td>DAL</td>
<td>-0.016</td>
</tr>
<tr>
<td>29</td>
<td>CHI</td>
<td>-0.017</td>
<td>NYK</td>
<td>-0.025**</td>
<td>IND</td>
<td>-0.017</td>
</tr>
<tr>
<td>30</td>
<td>MIA</td>
<td>-0.019*</td>
<td>BOS</td>
<td>-0.026**</td>
<td>DEN</td>
<td>-0.028**</td>
</tr>
</tbody>
</table>

* - significant at the 1% level. ** - significant at the 5% level.
*** - significant at the 10% level.

sampled time periods (about twice the sample average for a team-season) with an average duration of 1.8 minutes during the season and b) did not win many games overall. Over short spans, player, time-period level wins produced can be highly variable and sometimes quite small (non-pivotal) due to many zeros in the box score. This is reflected in the figure, where
there are modes at “wins produced” of about −0.02, 0 and 0.02. During this season, Houston and San Antonio were tied for the highest average wins produced of 0.0071. Were Utah able to increase its average per period “wins produced” by 0.0022 to that of Houston, Utah’s team chemistry would produce a slightly larger increase of $0.0022 \times 1.045 = 0.0023$ to a wins produced of 0.0072, ceteris paribus. In Table 2.1 most team-seasons (64 of 90) exhibit positive estimated peer-effects, but negative peer-effects occur. Median team peer-effects for each season are .015, .002, and .014. Significantly positive (negative) peer effects occur for 19 (4) of 90 team-seasons (at the 0.05 significance level). Peers may inhibit each other’s marginal win product in the present environment. However, most of the teams do not exhibit significant team chemistry. In 89 of 90 team-seasons, peer-effect estimates are smaller than those of Horrace, Liu, and Patacchini (2016) in their analysis of Syracuse University Men’s college basketball. This prevailing difference between sample peer-effects could be due to random variation, difference of environment, or specification of competitor-effects within the present study. If it is due to competitor effects, then it seems that ignoring them causes the peer-effect to be biased.

Despite roster turnover, several teams exhibit persistence in peer-effect across seasons. Minnesota held a significantly positive peer-effect for each of the three sampled seasons. Boston, San Antonio, Utah, and Phoenix held significantly positive peer-effects in two seasons. Based
on Spearman’s rho, the Western Conference exhibits greater peer-effects persistence from season-to-season. If \( \rho_{ij} \) is the rank correlation of peer-effects between season \( i \) and season \( j \) for all teams in a conference, then the Western conference had significant (at the 95% level) correlations of \( \rho_{12} = \rho_{23} = 0.404 \), while the same statistics for the Easter conference were \( \rho_{12} = 0.332 \) and \( \rho_{23} = -0.289 \), significant at the 95% and 90% levels, respectively.

As Schrage (2014) suggests, team peer-effects are the “holy grail” of productivity analysis in sports. As we have observed, the measurement of peer-effects requires accounting for player productivity at the period level, while simultaneously controlling for the contributions of other players. Explaining variation in peer-effects across team-season also presents a challenge. Oliver (2004) develops a four-factor model of average scoring differential (between a team and their opponents). In this model the factors most highly correlated with average scoring differential are a team’s shooting efficiency differential (against opponents), turnover rate differential, rebounding rate differential, and free throw rate differential. The model is highly explanatory of average score margin differentials in the NBA, while maintaining very low levels of right hand side variable dependence (Oliver, 2004). We calculated each of these four factors for each team-season in the data, and found that there is a negative and significant relationship between our estimated peer-effects and the free throw rate differential for each team, conditional on the other factors and team fixed-effects.\(^{22}\) Even in our moderately-sized sample, there is evidence that a higher free throw rate differential is a moderately strong indicator of weaker peer-effects. What is the potential mechanism generating this relationship? Teams that generate a high free throw rate tend to shoot more contested shots (e.g., isolation-penetration or post entry shots). A high rate of contested shots is often an indication that a team does not rely heavily on rapid, frequent ball movement (e.g., “weak side reversals” and “penetration kicks to the corner”) to generate open shots. Ball movement is intended (by its very nature) to generate peer-effects by allocating the ball away from the defense and toward open (relatively high-percentage) shots.

\(^{22}\)Results available by request.
2.4 Conclusion

We estimate peer-effects using a network production model that controls for competitor-effects, strategy and overall team play and find evidence of (generally positive) peer-effects (team chemistry). In so doing, we generalize the work of Horrace, Liu, and Patachini (2016), who develop a network production function model that estimates peer (teammate) effects upon player productivity in college basketball. Their peer-effects do not condition upon competitor effects, introducing omitted variable bias. Generalizing their model, we account for strategic decisions and contemporaneous play of opposing team via a competitor network. We apply this model to a three-year sample of NBA regular season game data. We estimate NBA team peer-effects conditioning on both teams’ strategies and abilities, rendering a more reliable team-chemistry measure. While generally positive, we find examples of significantly negative peer-effects among NBA teams. We also find that team chemistry tends to persist from year-to-year and that a team’s peer-effect is negatively related to its free throw rate differential (which is one of the four primary factors of score margin production in basketball), and we conjecture that teams with higher free throw rate differentials tend to take more contested shoots and are less likely to “find the open man.” Based on this, it may be interesting to develop peer-effect weighting schemes based on passing and ball sharing. To do this we could develop statistics based on how often two individual players pass the ball to one another. However, this is left for future research.
References


Chapter 3

Adaptive LASSO for Stochastic Frontier
Models with Many Efficient Firms

The LASSO (Tibshirani, 1996) is applied to select a subset of maximally efficient firms in the fixed-effect stochastic frontier model for panel data of Schmidt and Sickles (1981). Asymptotic properties of the estimator are derived in this context. Under regularity conditions the LASSO estimator exhibits the oracle property, and simulations suggest that it outperforms the least squares dummy variable estimator in terms of the root mean squared error of the estimated firm-level inefficiencies. An application of the LASSO to rice farm data suggests that the resulting subset of maximally efficient firms is comparable to the efficiency subsets calculated from the same data in Horrace and Schmidt (2000).
3.1 Introduction

Current estimators of the stochastic frontier (SF) model yield point estimates of firm-level efficiency, which (when ranked) imply that a single firm in the sample is most efficient. That is, SF model estimators do not allow for efficiency ties, yet there may be several firms in the sample tied for most efficient, and we would like to develop techniques to allow for this scenario. Current approaches adopt two-step methodologies to identifying a subset of efficient firms. In the first step, firm-level efficiencies (or equivalent measures) are estimated, and in the second step an inference technique or selection criterion is used to determine membership in a subset of most efficient firms. For example, in the parametric SF model of Aigner, Lovell and Schmidt (1977), there have been several papers to construct parametric prediction intervals for the conditional mean efficiency estimates based on Jondrow, Lovell, Materov and Schmidt (JLMS, 1982). Horrace and Schmidt (1996), Simar and Wilson (2010), and Wheat, Greene and Smith (2014) estimate JLMS efficiency and then construct univariate intervals that imply statistical indistinguishability of firms with the largest estimates. Horrace (2005a) and Flores-Lagunes, Horrace, and Schnier (2007) extend this to multivariate intervals that account for the multiplicity inherent in the ranked estimates. Using these intervals, they develop selection procedures that produce a subset of most efficient firms at a pre-specified error rate. Horrace and Schmidt (2000) develop multivariate intervals for the semi-parametric SF model of Schmidt and Sickles (1984) for panel data. Despite the semi-parametric model, their inference technique relies on a parametric assumption on the distribution of estimated efficiencies.

More recently, Kumbhakar, Parmeter and Tsionas (2013) propose a zero inefficiency stochastic frontier (ZISF) model for cross sectional data that produces a subset of firms in the sample that are fully efficient. They estimate the probability of a firm falling into the zero inefficiency regime using a latent class model, then use the probability to adjust (shrink) the individual
inefficiency estimates to reflect the presence of both efficient and inefficient firms in the sample. Using the parameter estimates, they compute individual posterior estimates of the probability of being fully efficient, and then, with a pre-specified cut-off, they assign each firm to the fully efficient regime or the inefficient regime. Unfortunately, Rho and Schmidt (2015) discuss an identification issue in this model. They point out that if there is little inefficiency in the sampled firms, it is hard to distinguish whether this is due to small variance of individual efficiencies or due to a large proportion of fully efficient firms, which leads to an observational equivalence in the likelihood function and a lack of identification of the model. Moreover, the ZISF model suffers from the same issues as the previously mentioned techniques; 1) it is parametric and 2) it is a 2-step procedure. We would like to develop models that are semi-parametric and identify a subset of efficient firms in a single step.

We propose a new one-step, semi-parametric procedure for identifying latent membership in a subset of efficient firms using LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996). Specifically, we develop estimation procedures which identify a subset of marginal effects and firm-level inefficiencies as exactly zero. The proposed adaptive LASSO estimation proceeds as least squared dummy variable (LSDV) estimation, but the object function is augmented with two penalty terms: the adaptively weighted shrinkage $L_1$ penalties for the input coefficients and for the firm-level inefficiencies. Here, we estimate and penalize the inefficiency terms directly. However, we show that this is equivalent to estimating the firm-level fixed effects and then penalizing their differences from the firm with the largest effect in the sample. We also show the object function can be equivalently solved by ‘within’

---

1The LASSO has received much attention in the statistics literature. For example, Wang et al. (2007a) consider its application to autoregression, Caner (2009) considers generalized methods of moments, Zhu et al. (2010) consider its application to spatial autoregression, and Lee et al. (2016) consider the selection of break points in time series. Recently in economics, the shrinkage technique is used in select valid and relevant moments: Belloni et al (2012), Cheng and Liao (2015), and Caner et al (2016) among others.

2Lu and Su (2016) use the adaptive group LASSO technique to select regressors and factors in the regression context. If our fixed effects model can be regarded as a special case of a factor error structure model, then our model and the model in Lu and Su (2016) are closely related. Cheng et al. (2016) use the LASSO to identify a break in the factor model where they selected a group of factor loadings generated due to the break in factors, but their model is a pure factor model, and they uses the group LASSO method to estimate the model.
transformed versions of estimation. The within transformation eliminates the time-invariant inefficiencies (and their associate fixed-effects), leading to a two-step revision of the LASSO estimation.

Asymptotics are for the case \((N, T) \to \infty\), where \(N\) is the number of firms and \(T\) is the number of time periods in the sample. We show that the proposed estimator possess the oracle property under regularity conditions. More precisely, the LASSO consistently selects model coefficients as \((N, T) \to \infty\). Conditional on selection consistency, the estimates for the selected marginal effects and the individual firm inefficiencies are \(\sqrt{NT}\)- and \(\sqrt{T}\)-consistent (respectively) in our setting whereas the estimates from standard estimation procedures, like LSDV, for common estimates are \(\sqrt{NT}\)- and \(\sqrt{T}/(\log N)^2\)-consistent (respectively) in the same setting. The efficiency difference in the estimators is more pronounced in the estimation of the common intercept (the maximum of the individual fixed effects): The LASSO estimator shows \(\sqrt{\delta_0 NT}\) consistency, where \(\delta_0\) is the fixed proportion of fully efficient firms, while the LSDV estimator exhibits \(\sqrt{T}/(\log N)^2\) consistency. Therefore, the LASSO estimator of the common intercept converges faster. Consequently, the LASSO may outperform LSDV (in terms of root mean squared error of the estimated common intercept) even when \(T\) is small, and this is borne out in our simulation study.

We apply the LASSO to Indonesian rice farm data previously analyzed by Erwidodo (1990), and Horrace and Schmidt (1996, 2000) among others. The LASSO selects a subset of maximally efficient rice farms that is comparable in size and composition to the Gupta subset of Horrace.

---

3The LASSO estimate for the intercept is estimated as a common intercept of the firms categorized as fully efficient by LASSO technique, so if we know the true model, we can use the \(\delta_0 NT\) observations to estimate the intercept (\(\delta_0\) is a fixed parameter so it is not relevant when calculating convergence rates. We intentionally leave it in the expression of the convergence rate for the common intercept to stress out the role of \(\delta\) in our estimation. In this case, we may redefine \(\delta_0 N = \delta_0\) and think \(\delta_0\) grows proportionally to \(N\).) However, the convergence rate of the estimate for the intercept in the standard fixed effect SF model can’t be established clearly because it relies on the maximum value of the estimated fixed effects to estimate the intercept and the rate the maximum value converges to the true intercept should be determined by the assumption for the distribution of the inefficiency. Park et al (1998) studies the asymptotic properties of the estimator for the common intercept in LSDV and shows that it has the convergence rate of \(O_p\left(\frac{\log N}{\sqrt{T}} + \frac{1}{N}\right)\) if a inefficiency has a shifted half normal or exponential distribution.
and Schmidt (2000). However, the LASSO does so without multivariate inference on the efficiency estimates of 171 rice farms and without the distributional assumptions that it entails.

The technique developed for SF models may be more broadly applicable. Whenever we have a linear regression model with individual fixed effects, and the ranked fixed effects contain important information, the LASSO may be applied to produce a subset of the best (or worst) effects. For example, consider an education outcome function. After controlling for other factors (i.e. family background and teacher quality), we may want to calculate a group of the best or worst students based on their individual-specific outcomes. Mutual fund performance may be another example. Practitioners may conveniently use the LASSO to calculate subsets of the best funds, based on average returns. Moreover, this type of "best and the rest" classification may be more important as dataset sizes grow (big data), because group-level classification may be more useful than individual ranking, when the number of comparisons is large.

The rest of this paper is organized as follows. The next section introduces the model and the adaptive LASSO estimator. Section 3 provides some technical assumptions and derives the oracle property of the estimator. Section 4 discusses computational issues such as tuning parameter selection and optimization algorithm. Section 5 and 6 provide simulation and empirical application results, and section 7 concludes. All the proofs are given in the Appendix.
3.2 Adaptive LASSO for many efficient firms

3.2.1 Production Function

The fixed effect SF panel data model with time-invariant technical inefficiency due to Schmidt and Sickles (1984) is,

\[ y_{it} = \alpha_0 + x_{it}' \beta_0 + \nu_{it} - u_{0,i} \text{ for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T \] (3.1)

where \( y_{it} \) is the logarithm of scalar output of the \( i^{th} \) firm in the \( t^{th} \) period, \( \alpha_0 \) is a scalar common intercept (common to all firms \( i \)), \( x_{it} \) is a \( p \times 1 \) input vector, \( \beta_0 \) is a \( p \times 1 \) corresponding parameter vector of marginal effects, and \( \nu_{it} \) is a two sided noise with \( \nu_{it} \sim iid(0, \sigma^2_{0,\nu}) \) and \( \nu_{it} \perp x_{it} \). The \( u_{0,i} \) are time-invariant firm-specific inefficiencies, which are treated as fixed effects. We assume \( u_i \geq 0 \) and independent of \( \nu_{it} \) and the inputs, but do not impose a distributional assumption on the inefficiency distribution. In matrix form the model is

\[ Y = \alpha_0 l_{NT} + X_1 \beta_0 - X_2 U_0 + \nu \] (3.2)

where \( l_{NT} \) is \( NT \times 1 \) vector with ones, \( X_1 = [x_{it}'] \), and \( X_2 = l_N \otimes l_T \).

Standard LSDV estimation (or equivalent within estimation) proceeds as follows. Rewrite equation (3.1) as \( y_{it} = \alpha_{0,i} + x_{it}' \beta_0 + \nu_{it} \), where \( \alpha_{0,i} = \alpha_0 - u_{0,i} \) are firm-specific fixed-effects. If \([X_1, X_2]\) is full column rank, then all the parameters of the model are identified, and we regress \( Y \) on \( X_1 \) and \( X_2 \) to get ordinary least squares (OLS) estimates \( \hat{\beta}_0 \) and \( \hat{\alpha}_{0,i} \), respectively. The OLS estimates are consistent for \( \beta_0 \) (as \( N \) or \( T \to \infty \)) and \( \alpha_{0,i} \) (as \( T \to \infty \)), respectively. Also, \( \hat{\alpha}_0 = max_j \hat{\alpha}_{0,j} \) is consistent for the common intercept as \( N \) and \( T \to \infty \) because \( min_j u_{0,j} \to 0 \) and \( max_j \alpha_{0,j} \to \alpha \) as \( N \to \infty \) as long as the data generating process for \( u_{0,i} \) allows \( u \) arbitrarily close to zero with positive density (Greene, 1980; Schmidt and Sickles, 1984).
The individual firm inefficiencies are accordingly consistently estimated by  \( \hat{u}_{0,i} = \hat{\alpha}_0 - \hat{\alpha}_{0,i} \) (as \( N \) and \( T \to \infty \)). In this case, \( \hat{\alpha}_0 \) represents maximal output in the population, and the individual \( \hat{u}_{0,i} \) are interpretable as absolute inefficiencies.

In practice, there are many reasons why \( [X_1, X_2] \) may not be full column rank, but a leading case is when \( X_1 \) contains time-invariant regressors. If so, the marginal effects of time-invariant regressors and the individual fixed-effects (and also the common intercept) are indecomposable within the model, which leads to a fundamental identification problem. See Greene (2005) and Feng and Horrace (2007) for detailed discussions about this issue and potential solutions.

Another interesting case to consider is when \( X_2 \) contains indicator or categorical variables that vary over \( t \). If so, the point estimate of \( \alpha_0 \) will vary with the omitted reference groups of the categorical variables, but the individual inefficiency estimates will not. In this case the estimated common intercept still can serve as an instrument to identify the individual firm-level inefficiencies, but the \( \hat{u}_{0,i} \) has to be interpreted as relative efficiencies. In general, it would take fortuitous circumstances for \( \alpha_0 \) to be identified, so \( \hat{u}_{0,i} \) is almost always interpreted as relative efficiency.

For the LASSO version of the fixed effect SF model in equation (3.1), we impose the following sparsity assumption on \( \theta_0 = (\alpha_0, \beta_0', U_0')' \).

**Sparsity Assumption** (i) \( \beta_0 = (\beta_{0,A}', \beta_{0,A_c}')' \) and \( n(\beta_{0,A}) = p_0 < p = n(\beta_0) \) where \( n(M) \) is the number of elements in \( M \), \( A = \{j : \theta_{0,j} \neq 0\} \) represents the index set for the nonzero coefficients in \( \theta_0 \), and \( A^c = \{j : \theta_{0,j} = 0\} \) is defined similarly so that \( \beta_{0,A} \) represents the true non-zero input coefficients. (ii) Similarly, \( U_0 = (U_{0,A}', U_{0,A_c}')' \) and \( n(U_{0,A}) < N \). Denote \( \delta_0 = \frac{n(U_{0,A_c})}{N} \), then the assumption is equivalently stated by \( \delta_0 > 0 \). By construction, \( \min_{j \in A} |u_{0,j}| = \eta > 0 \), however, it is allowed \( \eta \to 0 \) as \( (N, T) \to \infty \).

---

4 This is related to the identification issue in wage gap decomposition, discussed in Horrace and Oaxaca (2001). Their concern is identification of the wage gaps across various labor markets, while our concern is identification of efficiencies across firms.

5 For analytic simplicity, without loss of generality, it is assumed the last \( \delta_0 \times N \) of firms are fully efficient.
The sparsity assumption is common in the LASSO literature and implies only a subset of the regressors are relevant to the true model, which justify the use of penalized technique to recover the true model. In addition to the sparsity in the regressors as in the literature, we assume there’s a sparsity in inefficiency. For practical purposes, the assumption will be met when the population under analysis contains many highly efficient firms. This model becomes the standard fixed effect SF model if \( p_0 = p \) and \( \delta_0 = 0 \). It becomes the neo-classic production model, which assume every firm is efficient, if \( p_0 = p \) and \( \delta_0 = 1 \). We allow \( \eta \to 0 \) as \( (N, T) \to \infty \) as we do not restrict the lower bound of inefficiency in SF models.

### 3.2.2 Adaptive LASSO estimator

The LASSO version of the fixed effect SF model can be estimated by either the “within estimation” or the Least Squares Dummy Variable (LSDV) estimation. We start with the LSDV and then show it is equivalent to ‘within’ transformed version of estimation with a two-step procedures.

The adaptive LASSO estimator for \( \theta_0 \) is defined as

\[
\hat{\theta}(\Lambda, \Pi) = [\hat{\alpha}(\Pi), \hat{\beta}(\Lambda)', \hat{U}(\Pi)']
\]

\[
= \arg\min_{\alpha, \beta, U} \left\{ \sum_T \sum_N \{y_{it} - \alpha - x_{it}'\beta + u_i\}^2 + \Lambda \sum_{j=1}^p \hat{\lambda}_j|\beta_j| + \Pi \sum_{k=1}^N \hat{\pi}_k|u_k| \right\}
\]

\[
= \arg\min_{\theta} \left\{ (Y - X\theta)'(Y - X\theta) + \Lambda \sum_{j=1}^p \hat{\lambda}_j|\beta_j| + \Pi \sum_{k=1}^N \hat{\pi}_k|u_k| \right\} \quad \text{where } u_i \geq 0
\]

where \( X = [l_{NT}, X_1, -X_2] \), and \( \Lambda \) and \( \Pi \) are positive tuning parameters for \( \beta \) and \( U \), respectively. \( \{\hat{\lambda}_j\}_{j=1}^p \) are some data-dependent weights for \( \beta \), which are usually obtained from

\(^6\text{Note that } \hat{u}_{0,i} \text{ is estimated from } \hat{u}_{0,i} = \hat{\alpha}_0 - \hat{\alpha}_{0,i} \text{ in the standard SF model where } \hat{\alpha}_0 = \max_{j=1}^N \hat{\alpha}_{0,j}. \text{ We always have one relatively } 100 \% \text{ efficient firm from the model, however, it does not mean that it is zero inefficiency based on a absolute standard. It only become absolute zero inefficiency when } N \to \infty. \)
the absolute value of some consistent estimate. In this paper, we set \( \hat{\lambda}_j = |\beta_{j, LSDV}^{LSVD} - \gamma\beta| \) and \( \hat{\pi}_k = |\hat{u}_{k, LSDV}^{LSVD} - \gamma u| \) with some \( \gamma\beta > 0 \) and \( \gamma u > 0 \). \(^7\)

There are two things to be noted in (3.3). First, we are estimating \( \alpha \) and \( U \) in one step, which is not feasible in the standard fixed effect SF model because of identification problem (theoretically) or perfect multicollinearity (in practice). This is feasible in this model due to the presence of efficient firms, which allow us to identify \( \alpha \) and \( U \) separately while avoiding the multicollinearity problem. Second, it is also notable that we are using two different tuning parameters, and \( \{\beta_{j, LSDV}^{LSVD}\}_{j=1}^p \) and \( \{\hat{u}_{k, LSDV}^{LSVD}\}_{k=1}^N \) that have different convergence rates. That is, the former is a \( \sqrt{NT} \)-consistent estimate of \( \beta_{0,j} \) while the latter is a \( \sqrt{T/(\log N)^2} \)-consistent estimate of \( u_{0,k} \), which will be formally proved in lemma 3.4.1. We use the two tuning parameters because of the difference in asymptotic behaviors of the two estimator, \( \beta(\Lambda) \) and \( \hat{U}(\Pi) \), which will be discussed in details in the assumption 3.4.2.

**Remark 3.2.1** Our model is related to latent group structure models, in particular, the model in Su et. al. (2016). In Su et. al. (2016), the regression coefficients are heterogeneous across groups but homogeneous within a group, and group membership is unknown. Their methodology forces some of individual coefficients to have the same value by penalizing their difference from a group-specific coefficient value which is simultaneously estimated within their model. \(^8\) We can show our model has the same features by reparameterizing (3.1) as

\[
y_{it} = \alpha_{0,i} + x_{it}'\beta_0 + \nu_{it} \quad \text{where} \quad \alpha_{0,i} = \alpha_0 - u_{0,i} \quad \text{and} \quad \alpha_0 \geq \alpha_{0,i}.
\]

Then, using a similar penalized

\[^7\] It should be noted that \( \{\beta_{j, LSDV}^{LSVD}\}_{j=1}^p \) are from a preliminary estimation but \( \{\hat{u}_{k, LSDV}^{LSVD}\}_{k=1}^N \) may be obtained from the residuals after the first step estimation of (3.5).

\[^8\] Their methodology (also, our methodology) is related to the fused LASSO proposed by Tibshirani, Saunders, Rosset, Zhu, and Knight (2005) where the parameters of interest have an order in some meaning way and some parameters take the same value with the neighboring parameters. The fused LASSO encourages sparsity of the differences between the neighboring parameters by penalizing the differences. This technique is in particular useful when there are multiple changes in the parameter values along the natural order of the parameters (e.g. detecting multiple structural changes in the time series setting; Harchaoui and Lévy-Leduc (2010), Chan, Yau, and Zhang (2014), and Qian and Su (2015)).
technique, the reparameterized model can be estimated by

\[
[\hat{\alpha}(\Pi), \hat{\alpha}_I(\Pi), \hat{\beta}(\Lambda)] = \text{argmin}_{\alpha, \alpha_I, \beta} \left\{ \sum_T \sum_N \{y_{it} - \alpha_i - \chi_{it}'\beta\}^2 + \Lambda \sum_{j=1}^p \hat{\lambda}_j|\beta_j| + \Pi \sum_{k=1}^N \hat{\pi}_k|\alpha - \alpha_i| \right\}
\]

(3.4)

where \( \alpha \geq \alpha_i, \hat{\alpha}_I(\Pi) = [\hat{\alpha}_1(\Pi), \ldots, \hat{\alpha}_N(\Pi)] \). (3.4) is not much different from (3.3) conceptually and computationally, as the inefficiencies in the fixed effect SF model are nothing but the distances from the largest fixed effect in the sample.\(^9\) We are penalizing the differences between the largest fixed effect and the firm fixed effects in (3.4) to identify the subset of best firms, which implies our model has a classification feature similar to the one in Su et al. (2016).\(^10\) In the same spirit, we can modify the sparsity assumption on inefficiencies: For \( \alpha_0 \geq \alpha_{i,0}, \alpha_{i,0} = \alpha_0 \) if \( i \in BS_0 \) where \( BS_0 \) is the true set of efficient firms who have \( \alpha_0 \) (the leading group specific fixed effect) as for their fixed effects and \( BS_0 \neq \). In this regard, our model can be viewed as a technique for simultaneous classification and estimation of the firm-level efficiency.

With some algebra,\(^11\) we can concentrate out the \( \alpha \) and \( U \) in (3.3), which implies the above problem can be equivalently solved in two steps: In the first step solve

\[
\hat{\beta}(\Lambda) = \text{argmin}_{\beta} \left\{ \sum_T \sum_N \{y_{it}^* - \chi_{it}'^*\beta\}^2 + \Lambda \sum_{j=1}^p \hat{\lambda}_j|\beta_j| \right\}
\]

(3.5)

\[
= \text{argmin}_{\beta} \left\{ (Y - X_1\beta)'Q(Y - X_1\beta) + \Lambda \sum_{j=1}^p \hat{\lambda}_j|\beta_j| \right\}
\]

---

\(^9\)We may show they possess the same asymptotic properties using similar arguments in our asymptotic analysis. We verified that they produces similar estimation results in finite sample simulations (the differences in the estimation results from the two object functions were less than \(10^{-3}\) in many cases).

\(^10\)There are two distinct differences between our model and theirs. First, our group membership is determined by firm fixed effects (or firm-level inefficiencies) whereas it is structural parameters that determine membership in their paper. Bonhomme and Maurea (2015) consider a latent group structure problem, where group membership is determined by group specific fixed effects, but their methodology relies on minimization of a least squares criterion with respect to all possible groupings without the LASSO technique. Moreover, they don’t require our constraint on the fixed effects. Second, we are not estimating an arbitrary latent group structure, but we are identifying a “best and the rest” group structure by imposing the constraints \( \alpha \geq \alpha_i \).

\(^11\)We prove the equivalence between (3.3), and (3.5) - (3.6) in Appendix C.
where \( y^*_t = y_{it} - \bar{y}_i \), \( x^*_t = x_{it} - \bar{x}_i \), and \( Q \) is a within transformation matrix such that \( I_N \otimes (I_T - \frac{1}{T} i_T' i_T) \). Then the second step is

\[
[\hat{\alpha}(\Pi|\hat{\beta}(\Lambda)), \hat{U}(\Pi|\hat{\beta}(\Lambda))]'
\]

\[
= \text{argmin}_{\alpha, U} \left\{ \sum_T \sum_N \left\{ y_{it} - \alpha - x_{it}' \hat{\beta}(\Lambda) + u_i \right\}^2 + \prod_k \sum N \hat{\pi}_k |u_k| \right\} \quad \text{where } u_i \geq 0
\]

\[
= \text{argmin}_{\alpha, U} \left\{ (Y - X\theta(L))' (Y - X\theta(L)) + \prod_k \sum N \hat{\pi}_k |u_k| \right\}
\]

where \( \theta(L) = [\alpha, \hat{\beta}(\Lambda), U']' \)\(^{12} \) For notational simplicity, we denote \( \hat{\alpha}(\Pi|\hat{\beta}(\Lambda)) \equiv \hat{\alpha}(\Pi) \) and \( \hat{U}(\Pi|\hat{\beta}(\Lambda)) \equiv \hat{U}(\Pi) \). Figure 1 in chapter 3 of Hastie et al (2009) visually explains how the \( L_1 \) penalty often leads to zero estimates for some parameters. The LASSO shrinks the estimates toward 0, as the tuning parameter increases, and leads to exactly zero values for some parameters, because of a singularity at the origin. The initial LASSO by Tibshirani (1996) has no weight on its penalty term, so it shrinks each estimate equally, which leads to a trade-off between consistent estimation and consistent variable selection (Fan and Li, 2001; Zou, 2006). To address this, Zou (2006) proposes the adaptive LASSO which puts different weights onto each penalty term for each parameter, enabling adjustment of the degree of shrinkage for each parameter, using the information from preliminary consistent estimation. He proves the adaptive LASSO possess the oracle property. In this paper, we set \( \hat{\lambda}_j = |\hat{\beta}_j^{LSDV}|^{-\gamma_3} \). If the true parameter is zero, the \( |\hat{\beta}_j^{LSDV}| \) would be close to zero as \( N \to \infty \) or \( T \to \infty \), which, in turn, leads to \( \hat{\lambda}_j \to \infty \), so we would be more likely to have a zero estimate for the parameter in this case. However, it is impossible to completely remove the bias in nonzero parameter estimates, so the tuning parameter should be large enough to select the zero parameters but not too large as to induce bias in the nonzero parameter estimates. Selection of the tuning parameters and their asymptotic conditions to achieve oracle property will be discussed in

\(^{12}\)As we consider a two step procedure, we may adopt the hybrid estimation approach of Efron et. al. (2004) and Hui et. al. (2015) to reduce the bias from using the shrinkage technique in the first step. That is, instead of \( \hat{\beta}(\Lambda) \), we compute the “within” estimate for \( \hat{\beta}_0 \) under the model selected in the first step and use the unpenalized estimate for the second step.
next sections.

3.3 Computation

3.3.1 Optimization algorithm

The $L_1$ penalty term in the object function has no second derivative at the origin, so we can’t directly apply standard quadratic optimization algorithms (e.g. Newton-Raphson). Many alternative optimization algorithms have been developed: Least Angle Regression (Efron et al., 2004), Local quadratic approximation (Fan and Li, 2001), Coordinate Descent Algorithm (Friedman et al, 2008), among others. Optimization of (3.5) is a standard LASSO problem, so we may use one of the algorithms for implementation of the first step.

In the second step, we wish to obtain a result that retains $\hat{u}_{0,j}(\Pi) \geq 0$ in order to be consistent with the model assumption of $u_{0,j} \geq 0$. Note that, if we impose a positive constraint on $U$ in optimization procedure, then (3.6) becomes a standard constrained quadratic optimization problem with no singularity. Therefore, we may use one of the standard constrained optimization algorithms (e.g. Sequential Quadratic Programming) for implementation of (3.6). However, this may be computationally costly because the number of constraints in our problem is $N$. Alternatively, we propose a coordinate decent algorithm which produces almost the same estimation results as the standard constrained optimization but without computational cost. Using preliminary inefficiency ranking information among the firms from the LSDV estimation, this algorithm allows us to skip a large number of irrelevant optimization steps. The algorithm is implemented as follows. For a simplicity, we suppress the $(\Pi)$ notation.

1. Using $\beta(\Lambda)$ from the first step, compute $\hat{\alpha}_i = \frac{1}{T} \sum_t y_{it} - x_{it}' \beta(\Lambda)$ and $\hat{u}_i = \max_{j=1}^N \hat{\alpha}_j - \hat{\alpha}_i$.

As mentioned earlier, we may use the “within” estimator for $\beta_0$ from the model selected in the first step.
for all $i$. Let $\hat{\alpha}[1] \leq \hat{\alpha}[2] \leq \ldots \leq \hat{\alpha}[N]$ be the rankings of the $\hat{\alpha}_i$, so $\hat{\alpha}[N] = \max_{j=1}^{N} \hat{\alpha}_j$. Similarly, let $\hat{u}[N] \leq \hat{u}[N-1] \leq \ldots \leq \hat{u}[1]$ be the rankings of the $\hat{u}_i$, so $\hat{u}[N] = \min_{j=1}^{N} \hat{u}_j$. We set the initial value for $\alpha$ to $\hat{\alpha}[N]$. Denote the current values for $\hat{u}[i]$ and $\hat{\alpha}$ as $\hat{u}_i^{(0)}$ and $\hat{\alpha}^{(0)}$. Note that as we set $\hat{\alpha}^{(0)} = \hat{\alpha}[N]$, we have one fully efficient firm, $\hat{u}_i^{(0)} = 0$, now.

2. For a given $\Pi$, sequentially check the KKT condition for the second best firm, the third best firm,... That is, check the sign of $\Delta_{[N-i]} = \hat{u}_{[N-i]}^{(0)} - \prod_{k=N-i}^{i} \hat{\pi}_{[N-k]}^{T}$ from $i = N - 1$ to $1$.

(a) IF $\Delta_{[N-i]} \leq 0$, update $\hat{u}_{[N-i]}^{(0)}$ as $\hat{u}_{[N-i]} = 0$, and update $\hat{\alpha}$ as $\hat{\alpha} = \frac{1}{i+1} \sum_{k=0}^{i} \hat{\alpha}_{[N-i]}$. As we have new $\hat{\alpha}$ (the frontier parameter), we update the rest of the inefficiencies (from $[N-1-i]$ to $[1]$) as $\hat{u}_{[N-i-j]} = \hat{u}_{[N-i-j]}^{(0)} - (\hat{\alpha}^{(0)} - \hat{\alpha})$ for $j = 1, \ldots, N - i - 1$.

Then, go back to 2 and check next firm’s KKT condition.

(b) IF $\Delta_{[N-i]} > 0$, update $\{\hat{u}_{[N-i]}, \ldots, \hat{u}_{[1]}\}$ as $\hat{u}_{[N-k]} = \hat{u}_{[N-k]}^{(0)} - \prod_{k=N-k}^{i} \hat{\pi}_k^{T}$ for $k = i, \ldots, N - 1$. Repeat below LOOP until the absolute difference in the estimation results in two consecutive steps is smaller than a pre-specified threshold and then report the results.

LOOP

i. Update $\hat{\alpha}$ as $\hat{\alpha} = \hat{\alpha}^{(0)} - \frac{1}{N} \sum_{k=i}^{N-1} (\hat{u}_{[N-k]}^{(0)} - \hat{u}_{[N-k]})$

ii. Update $\{\hat{u}_{[N-i]}, \ldots, \hat{u}_{[1]}\}$ as $\hat{u}_{[N-k]} = \left(\hat{u}_{[N-k]}^{(0)} - (\hat{\alpha}^{(0)} - \hat{\alpha})\right)_+$ for $k = i, \ldots, N - 1$, where $(x)_+ = x$ if $x \geq 0$ and $= 0$ otherwise.

We provide a figure in Appendix C illustrating the above procedure. This coordinate decent algorithm uses the convexity of the object function and the preliminary inefficiency ranking at the same time, enabling us to reach the minimum of the object function quickly. We compare the series of estimation results between this algorithm and the Sequential Quadratic Programming (SQP) algorithm in Matlab in Appendix C. We find that the two algorithms
generally produce similar results, but the new algorithm is much faster than SQP.\footnote{For one replication with a sample size \((N, T) = (20, 10)\), the new algorithm took on average 1.5 second whereas the standard algorithm took 345 seconds. This gap will be pronounced as \(N\) increases.}

In the algorithm, the LOOP is necessary to optimize the object function, however, we can see that it shrinks \(\hat{\alpha}(\Pi)\) as well. This is not desirable because it may slow down the convergence rate of \(\hat{\alpha}(\Pi)\) and, in turn, it may induce bias on \(\hat{U}(\Pi)\) when \(T\) is small. In order to prevent this, we intentionally skip the LOOP in the implementation of our algorithm.\footnote{This modification can be viewed as a bias correct procedure. The modified algorithm helps to achieve better asymptotic and simulation results, which we shall see in subsequent sections.} The asymptotic results are derived from this modified algorithm and we use it for the simulation study and empirical exercises.

### 3.3.2 Tuning parameter

The performance of the adaptive LASSO estimator relies on an appropriate selection of the tuning parameters, and the CV and AIC criteria have been used in the LASSO literature. However, they lead to inconsistent model selection; too many nonzero estimates. Wang et al (2007b) shows that the tuning parameters based on a BIC-type criterion can identify the true model consistently. Therefore, in this paper, we consider the BIC type criteria for the selection of the two tuning parameters such that:

\[
(\Lambda^*, \Pi^*) = \operatorname{argmin}_{\Lambda, \Pi} \log \hat{\sigma}^2(\hat{\theta}(\Lambda, \Pi)) + \frac{|\hat{\beta}(\Lambda)| \log (NT) + |\hat{U}(\Pi)| \log (T)}{NT}
\]  

(3.7)

where \(\hat{\sigma}^2(\hat{\theta}(\Lambda, \Pi))\) is the mean squared error based on \(\Lambda\) and \(\Pi\). The criterion of Wang et al (2007b) is a special case with \(T = 1\). Equation (3.7) can be implemented using a two-dimensional grid search: 1) for every possible \(\Lambda\), implement the first step and compute

\footnote{We also experimented various types of selection criterions in the simulation study (e.g ERIC: Hui et al (2015) and \(I_{C_{p1}}\): Bai and Ng (2002)) and found (3.7) worked best in various panel structure. The two other criterions tended to select more sparse model than (3.7), however, the difference in the model selection between the criterions were not big.}
\( \hat{\beta}(\Lambda) \); 2) with the \( \hat{\beta}(\Lambda) \)s, implement the second step and compute \( \hat{U}(\Pi|\hat{\beta}(\Lambda)) \); 3) choose \( \Pi^* \) based on (3.7) for every \( \Lambda \); and 4) choose \( \Lambda^* \) based on (3.7) using the results from the previous step.\(^{17}\)

### 3.4 Asymptotic Theory

Asymptotics are for the case \((N, T) \to \infty\). Our analysis below builds on Zou (2006), and Zou and Zhang (2009), among others. Let LSDV estimates be denoted as \( \hat{\theta}(0) = [\hat{\alpha}(0), \hat{\beta}(0)', \hat{U}(0)']' \), where \( \hat{\alpha}(0) = \max_i \hat{\alpha}_i(0) \), \( \hat{\alpha}_i(0) \) are the LSDV estimates for individual intercepts, which is \( (X_2'X_2)^{-1}X_2(Y - X_1\hat{\beta}(0)) \), and \( \hat{u}_i(0) = \hat{\alpha}(0) - \hat{\alpha}_i(0) \). We now discuss consistency and the rate of convergence of the LSDV estimator in our setup. For this, we assume

**Assumption 3.4.1** For all \( i = 1, \ldots, N, t = 1, \ldots, T, j = 1, \ldots, P \) in the model matrix \( X \), \( \max_{i,t,j}|x_{it,j}| < \infty \) w.p.a 1. And, for any model \( \omega \) identified in the interval \([\Lambda_{\text{min}}, \Lambda_{\text{max}}]\) and \([\Pi_{\text{min}}, \Pi_{\text{max}}]\), we have \( c_1 \leq \text{eig}_{\text{min}}(\frac{1}{N^T}X_2'X_2) < \text{eig}_{\text{max}}(\frac{1}{N^T}X_2'X_2) \leq c_2 \) where \( c_1 \) and \( c_2 \) are some positive constants, and \( \text{eig}_{\text{min}}(\cdot) \) and \( \text{eig}_{\text{max}}(\cdot) \) denote minimum and maximum eigenvalues of some positive definite matrices, respectively.

Assumption 3.4.1 requires the regressor matrix to be well behaved. That is, the minimum eigenvalue grows by \( T \) whereas the maximum eigenvalue grows by \( NT \). The assumption implies \( c_1 \leq \text{eig}_{\text{min}}(\frac{1}{N^T}X_{1,\omega}'QX_{1,\omega}) \leq \text{eig}_{\text{max}}(\frac{1}{N^T}X_{1,\omega}'QX_{1,\omega}) \leq c_2 \) because \( X_{1,\omega}'QX_{1,\omega} \) is a

\(^{17}\)One may simplify the computation by using the criteria sequentially such that: For the selection of \( \Lambda \),

\[
\Lambda^* = \arg\min_{\Lambda} \log \hat{\sigma}^2(\hat{\beta}(\Lambda)) + |\hat{\beta}(\Lambda)| \frac{\log NT}{NT} \quad (3.8)
\]

where \( \hat{\sigma}^2(\hat{\beta}(\Lambda)) \)'s degree of freedom is \( N(T-1) \). And for the selection of \( \Pi \),

\[
\Pi^* = \arg\min_{\Pi} \log \hat{\sigma}^2(\hat{U}(\Pi)|\hat{\beta}) + |\hat{\beta}(\Lambda)| \frac{\log NT}{NT} + |\hat{U}(\Pi)| \frac{\log(T)}{NT} \quad (3.9)
\]

There may be a loss of precision from this method but it will reduce the computational cost significantly.
submatrix of $X'_ωX_ω$. Usually, the upper bounds of the estimation efficiency of the LASSO estimator is that of the baseline estimator based on the correct model. However, the adaptive LASSO estimator and the LSDV estimator in our setup exhibits different convergence rates for the common intercept and the inefficiency estimates, as will be shown below. First, we derive the convergence rate for LSDV estimator.

Lemma 3.4.1 Under Assumption 3.4.1, $E((\hat{\alpha}(0) - \alpha_0)^2) = O(\frac{(\log N)^2}{T})$, $E(||\hat{\beta}(0) - \beta_0||^2) = O(\frac{1}{NT})$, and $E((\hat{u}_i(0) - u_{0,i})^2) = O(\frac{(\log N)^2}{T})$ for $\forall i$ as $(N,T) \to \infty$.

Proof of this lemma and other lemmas and theorem is contained in Appendix C. Lemma 3.4.1 shows that each element in $\hat{U}(0)$ is $\sqrt{T}/(\log N)^2$ consistent. In the standard fixed effect SF model, it is not assumed that at least one firm in the sample is efficient, so it requires that $N \to \infty$, to ensure that we sample the first efficient firm. In our setup, we assume that we have at least one efficient firm, so practically speaking, only $T \to \infty$ is necessary. Next, for the oracle proof, we define

$$\hat{\theta}_A(\Lambda, \Pi) = \text{argmin}_\theta \left\{ (Y - X_A \cdot \theta)'(Y - X_A \cdot \theta) + \Lambda \sum_{j \in A} \hat{\lambda}_j |\beta_j| + \Pi \sum_{k \in A} \hat{\pi}_k |u_k| \right\}$$

(3.10)

where $X_A$ consists of columns of $X$ that correspond to the elements in $A$.

We derive the consistency and rate of convergence of $\hat{\theta}_A(\Lambda, \Pi)$ below. We require following assumptions.

Assumption 3.4.2 Denote $T^* = \frac{T}{(\log N)^2}$. (i) $\lim_{T,N \to \infty} \frac{\Lambda}{\sqrt{NT}} = 0$ and $\lim_{T,N \to \infty} \frac{\Lambda}{\sqrt{NT}}(NT)^{(\gamma_\beta)/2} = \infty$ with $\gamma_\beta > 0$ (ii) $\lim_{T,N \to \infty} \frac{\Pi}{\sqrt{T}^{*\gamma_u/2}} = 0$ and $\lim_{T,N \to \infty} \frac{\Pi}{\sqrt{T}^{*\gamma_u/2}} = \infty$ with $\gamma_u > 0$ (iii) $\beta_{0,A}$ is bound from below by $c >> 0$, and $U_{0,A}$ and $\lim_{T,N \to \infty} \left( \frac{\sqrt{T}}{\Pi} \right)^{1/\gamma_u} \eta = \infty$ where $\eta = \min(|u_{0,k}|)$

Similar assumptions can be found in Zou (2006), and Zou and Zhang (2009). Assumption 3.4.2 is crucial for the oracle property, because it controls the behavior of the tuning parameters, $\Lambda$ and $\Pi$, so they can select the zero coefficients properly without producing asymptotic bias.

---

18The two steps implementation can also be applied to this problem (this time, the $Q = Q_A$ where $Q_A$ is a within transformation matrix corresponding $X_{2,A}$)
in the nonzero coefficient estimates. Assumption 3.4.2 (ii) implies a condition that $T$ has to grow faster than $(\log N)^2$ so that $\frac{(\log N)^2}{T} \to 0$, however, the restriction is not strong as it covers many panel structure. Assumption 3.4.2 (iii) restricts the convergence speed of the nonzero coefficients to zero so that they can be distinguished from the zero coefficients by the estimation procedure. This is important in our context as we do not impose any lower bound for inefficiencies in the SF literature.

**Lemma 3.4.2** Under Assumption 3.4.1 and 3.4.2, $E((\hat{\alpha}(\Pi) - \alpha_0)^2) = O(\frac{1}{N^3 T})$, $E(||\hat{\beta}_A(\Lambda) - \beta_{0,A}||^2) = O(\frac{1}{N^3 T})$ and $E((\hat{u}_{A,i}(\Pi) - u_{0,A,i})^2) = O(\frac{1}{T})$ for $\forall i$ as $(N, T) \to \infty$.

This Lemma shows that if we select the correct model, the fixed effect LASSO estimator is an consistent estimator for $\theta_0$. The proof of this lemma in Appendix C shows that the mean square error of the LASSO estimator can be decomposed into two parts: the first part due to the penalty terms, and the second part due to the two-sided random error. Under these assumptions, the first part vanishes faster than the second, so estimation consistency is achieved.

From this lemma, we observe that $\hat{\alpha}(0)$ and $\hat{\alpha}(\Pi)$ have different convergence rates. LSDV uses only $T$ observation to estimate $\alpha_0$ and the max operator further slows down the rate (Park et. al., 1998). The $\hat{\alpha}(\Pi)$ is estimated as a common intercept of the firms categorized as efficient by the LASSO, so if we knew the true model, we could use the $\delta_0 N T$ observations to estimate $\alpha_0$. This result, in turn, leads to the convergence rate difference between $\hat{U}(0)$ and $\hat{U}_A(\Pi)$. However, we need to note that the difference in the convergence rates can only be observed after achieving selection consistency. That is, only when we can identify the true group of efficient firms will $\hat{\alpha}(\Pi)$ show a faster convergence rate. Otherwise, estimation error in $\hat{U}_A(\Pi)$ will be transfered into the estimation of $\alpha_0$, and the optimal convergence rate will not be achieved.

We will proceed to the oracle proof. First, we derive a useful lemma for the oracle of
\( \hat{\theta}(\Lambda, \Pi) \). Next lemma shows that \( \hat{\theta}(\Lambda, \Pi) \) estimates all the elements in \( A^c \) as zero w.p.a 1 as \((N, T) \to \infty \).

**Lemma 3.4.3** Under Assumption 3.4.1 and 3.4.2, \( \hat{\theta}(\Lambda, \Pi) \) estimates all the elements in \( A^c \) as zero w.p.a 1 as \((N, T) \to \infty \).

This lemma tells us that asymptotically \( \hat{\theta}(\Lambda, \Pi) \) works as if it knows the true model by estimating the zero coefficients exactly as zero. As in the proof of this lemma in the Appendix, the assumptions on \( \Lambda, \Pi, \) and \( \eta \) are crucial for this asymptotic characteristic of \( \hat{\theta}(\Lambda, \Pi) \).

**Assumption 3.4.3** \( \frac{x_{iA}'Q_{X_{1A}}}{NT} \to_p \sum_{A1}, \delta_0 \frac{i'x_{1A}}{\delta_0 NT} \left( \frac{x_{1A}'Q_{X_{1A}}}{NT}^{-1} x_{1A}' \right) \to_p \sum_{A} \), and the error terms and regressors satisfy the regularity conditions for Central Limit Theorem.

The next theorem confirms the oracle property of the adaptive LASSO estimator.

**Theorem (Oracle Property)** Under Assumption 3.4.1, 3.4.2 and 3.4.3, the LASSO estimator for \( \theta_0, \hat{\theta}(\Lambda, \Pi) \), has the oracle property. That is, the estimator satisfies:

1. Consistency in selection: \( Pr(\{j : \hat{\theta}(\Lambda, \Pi)_{j} \neq 0\} = A) \to 1 \) as \((N, T) \to \infty \)

2. Asymptotic normality
   
   1) \( \sqrt{NT}(\hat{\beta}_A(\Lambda) - \beta_{0,A}) \to_d N(0, \sigma_{v,0}^2 \Sigma_{A1}^{-1}) \)
   
   2) \( \sqrt{\delta_0 NT}(\hat{\alpha} - \alpha_0) \to_d N(0, \sigma_{v,0}^2 \Sigma_{\alpha}) \)
   
   3) \( \sqrt{T}(\hat{u}_i(\Pi) - u_{i,0}) \to_d N(0, \sigma_{v,0}^2) \)

w.p.a 1 as \((N, T) \to \infty \) where \( \Sigma_{\alpha} \) is from \( 1 + \delta_0 \frac{i'x_{1A}}{\delta_0 NT} \left( \frac{x_{1A}'Q_{X_{1A}}}{NT}^{-1} x_{1A}' \right) \to_p \sum_{A} \).

This theorem establishes the selection consistency and asymptotic normality of the adaptive LASSO estimator. In other words, asymptotically speaking, we can select the right model and identify a set of efficient firms without loss of estimation efficiency in this procedure.
One interesting thing is that the asymptotic distribution for the estimate of inefficiency is somewhat different from the one derived in Park et al (1998). Their results are based on the standard SF model and imply that the asymptotic variance of each inefficiency estimate is at least $2 \times \sigma^2_{v,0}$ and there will be more variation due to the uncertainty over whether one of the firm in the sample achieves full efficiency or not. However, in our model, the assumption of at least one fully efficient firm\(^{19}\) and the faster convergence rate of $\hat{\alpha}(\Pi)$ than that of $\hat{u}_i(\Pi)$ significantly reduce uncertainty in the estimation of $\hat{u}_i(\Pi)$.

3.5 Simulations

3.5.1 Setup

In this section, we study the finite sample performance of the estimator. We set up the model by defining $\alpha_0 = 1$, $\beta_0 = [\beta, \beta, \beta, 0, 0, 0, 0, 0]'$ with $\beta = 1$, $x_{it} \sim N(0, \Sigma)$ with the $(i,j)$-th element of $\Sigma$ set to $0.5|i-j|$, and $v_{it} \sim N(0, 1)$. We assume 30% of firms in the sample are fully efficient firms ($\delta_0 = 0.3$), and in every simulation each nonzero individual inefficiency is identically and independently generated from an exponential distribution $\frac{1}{\sigma_u} e^{-u_{it}/\sigma_u}$.\(^{20}\) We experiment with $\sigma_u \in \{1, 2, 4\}$. As $\sigma_u$ gets smaller, the selection problem becomes more difficult because the probability of small inefficiency draws will be high, making it more difficult for the LASSO to distinguish them from zero. This would be particularly difficult when the sample size is small as it is likely that Assumption 3.4.2 \((iii)\) in the asymptotic analysis is violated. Figure 3.1 shows the PDFs of inefficiency for each $\sigma_u$ value (left) and an example of draws from each PDF (right). We can clearly see that inefficiencies are densely packed near zero when $\sigma_u = 1$.

We set $\gamma_\beta = 1$ and $\gamma_U = 2$\(^{21}\) and the optimal tuning parameters are selected by (3.7) from a

---

\(^{19}\)This also allow us to derive a deterministic form of asymptotic distribution for the estimate of $\alpha_0$. 

\(^{20}\)We add an arbitrary small number (e.g. 0.01) to each inefficiency draw to ensure they are not zero.

\(^{21}\)We set $\gamma_U = 2$. However, we are free to choose the value of $\gamma_U$ as long as it is positive. From the asymptotic
two dimensional grid search over $10^{\text{linspace}(\log_{10}(10^{-4}/NT):1:50)} \times NT$ for $\Lambda$ and $10^{\text{linspace}(\log_{10}(10^{-4}/T):1:250)} \times T$ for $\Pi$ where $\text{linspace}(a:b:c)$ is a row vector of $c$ evenly spaced points between $a$ and $b$. We simulate each model 1,000 times with twelve combinations of $N \in \{100, 200, 1000\}$ and $T \in \{10, 30, 50, 70\}$.

3.5.2 Results

In what follows we only discuss results on $\hat{U}(\Pi)$, because it is our focus. Results on $\hat{\beta}(\Lambda)$ are in Appendix C. We report two types of statistics:

1. Estimation accuracy: a) square root of mean squared error (RMSE: $\sqrt{E(\hat{U}(\Pi) - U_0)^2}$), b) $\hat{\alpha}(\Pi)$ and $\hat{\alpha}(0)$, and c) Rank correlation between $U_{0,A}$ and $\hat{U}(\Pi)_A$.

2. Selection accuracy: a) $Pr_{U_A}$, b) $Pr_{U_{Ae}}$, c) $\delta_0$, and d) Max $\tilde{U}_A$.

analysis, we can see that setting a higher value for $\gamma_U$ ensures the LASSO estimates zero coefficients as zero, but also increases the probability of estimating (small) nonzero coefficients as zero. Therefore, in empirics $\gamma_U$ should be determined in light of this trade-off.

22We use a denser grid for $\Pi$ than that for $\Lambda$ because there are many inefficiency draws close to zero.
where $Pr_{U_A}$ is the probability of yielding nonzero estimates for $U_A$; $Pr_{U_{Ac}}$ is the probability of yielding zero estimates for $U_{Ac}$; $\hat{\delta}_0$ is the proportion of the firms estimated as efficient; Max $\bar{U}_A$ is the maximum of $U_A$ that are estimated as zero, that represents the worst case selection error when a model is underfitted. Table 3.1 and Figure 3.2 present the estimation accuracy results and Table 3.2 and Figure 3.3 present the selection accuracy results.

**Estimation Accuracy of $\hat{U}(\Pi)$:** Table 3.1 reports and compares the three types of estimation accuracy results from LASSO and LSDV. The results can be summarized as follows:

- As $T$ and $\sigma_u$ increase, the RMSE from the LASSO decreases, but the effect of $\sigma_u$ on RMSE is small. This is due to the fact that $\sigma_u$, which determines the frequency of near zero inefficiencies, significantly affects the selection performance as shown in Table 3.2. However, it may not be case for the RSME, because small inefficiency draws are already near zero, so that they do not contribute much to the RMSE.

- The LASSO outperforms LSDV in terms of RMSE. Differences in the estimation error among the zero inefficiency draws may be one explanation for the difference. However, the main explanation is the persistent overestimation of $\alpha_0$ in LSDV. Figure 3.2 presents the distribution of the estimates of $\alpha_0$ from the LASSO (solid line) and LSDV (dashed line). The distributions from the LASSO over the replications are centered close to the true value even when $T$ and $\sigma_u$ are small, and the variation in the distribution is decreasing significantly as $N$ or $T$ increase. However, those from LSDV are consistently displaced away from the true value. This is due to the fact that the LASSO estimator for $\alpha_0$ converges faster than the LSDV estimator as shown in the asymptotic analysis, and the max operator that LSDV uses to estimate $\alpha_0$ tends to pick up the most biased

---

23 $Pr_{U_A}$ is computed from averaging the percentage of nonzero estimates for $U_A$ in each replication. $Pr_{U_{Ac}}$ is computed from the same manner.

24 In the literature, underfitting means the case when we estimate one of the nonzero coefficients as zero.
individual intercept estimate ($\alpha_i$) of the zero inefficiency firms. In short, in the presence of a group of zero inefficiency firms, the max operator produces a biased estimate for $\alpha_0$, which, in turn, leads to a significant bias in the estimation of the inefficiencies in LSDV.

![Figure 3.2: Distribution of estimates for $\alpha_0$ from LASSO and LSDV](image)

The LASSO and LSDV show similar rank correlation results. We need to note that these results are computed only among the nonzero inefficient firms, and the LASSO achieves this even after selecting fully efficient firms, which implies the LASSO preserve the original ranking well when compared to LSDV.

**Selection accuracy of $\hat{U}(\Pi)$**: Table 3.2 present the results for $Pr_{Ua}$, $Pr_{Uac}$ and $\hat{\delta}_0$. For the reader’s convenience, we visualize the selection results in Figure 3.3 using the results from 12 cases with $N = 100$. The left panel is for $Pr_{Ua}$, and the right panel is for $Pr_{Uac}$, and the variables on each axises are given by $(x, y, z) = (\sigma_u, T, \text{probability})$. Note

Moreover, the LASSO shows slightly better results in many cases. The rank correlations are computed only from the nonzero inefficient firms, so the differences only come from the penalized effects. The difference is the biggest when $T$ and $\sigma_u$ are small, which is when we have a large uncertainty in the inefficiency estimates. This result may imply that the penalized technique may improve estimation of the nonzero inefficiencies as well, when LSDV is not reliable.
that the graph is drawn continuously but the results are not. The continuous planes are generated from connecting 12 coordinates and we add a color to each probability level to show the general trend of the selection performance depending on \((\sigma_u, T)\).

We can see that when \(T\) and \(\sigma_u\) are small, the LASSO incorrectly estimates many nonzero inefficiencies as zeros. However, the under fitting problem gets improved as \(T\) or \(\sigma_u\) increases. The under fitting or over fitting problem shouldn’t be exaggerated because most of the firms incorrectly estimated as zero inefficiency would have near zero inefficiency. The small values of Max \(\tilde{U}_A\) in Table 3.2 imply only the firms near the threshold of zero inefficiency may be incorrectly categorized as fully efficient. This may not be a serious problem in practice.

More importantly, it is impressive that even when \(T\) is small, including the \((N, T) = (1000, 10)\) case, the \(Pr_{U_A}\) and \(Pr_{U_A^c}\) are close to 1 if \(\sigma_u = 4\), which implies selection performance is more dependent on the distribution of the firms’ inefficiency than the size of \(T\) in finite samples. This gives us an important implication, that our model can be used in various panel structures, not limited to the case where \(T\) is large, as long as there are not too many near zero inefficiencies, and our primary interest lies in identification of fully efficient firms rather than individual inefficiency estimates.

\[\text{Figure 3.3: Visualization of the selection performance (N = 100)}\]
3.6 Empirical application

3.6.1 Comparison with "Ranking and Selection" (R&S) procedure

In this section, we apply our LASSO model to the rice farm data previously analyzed by Erwidodo (1990), Horrace and Schmidt (1996) and Horrace and Schmidt (2000), among others. In our context, the LASSO is designed to select a group of efficient firms, \( BS_0 \). The idea of selecting a subset of best firms is related to the R&S literature. R&S proceeds as follows. Suppose we estimate the LSDV model, and it yields \( \tilde{\alpha}_i \) for \( i = 1, \ldots, N \). R&S is an inferential decision rule that selects some subset of the populations that contain the population with largest (best) value of \( \alpha_i \) with some pre-specified error rate. If in truth \( \alpha_i = \alpha_0 - u_i \), this is equivalent to selecting populations with \( u_i \) closest to zero at the pre-specified error rate. If so, the connection between the LASSO and the R&S should be clear. The LASSO selects some subset of \( u_i \) to be zero in our model, while R&S select some subset of \( u_i \) closest to zero in a statistical sense. For more details of R&S procedure and its applications to economic problems, see Horrace and Schmidt (1996) and Horrace and Schmidt (2000).

To compare the two methodologies, we use data on 171 rice farms in Indonesia, observed for three wet and three dry seasons from six different villages. For a complete discussion of the data see Erwidodo (1990, unpublished manuscript). The empirical model is a standard Cobb-Douglas (loglinear) production function. Inputs to the production of rice included in the data set are seed (kg), urea (kg), trisodium phosphate (TSP) (kg), labor (labor-hours), and land (hectares). Output is measured in kilograms of rice. The data also include dummy variables. DP equals 1 if pesticides were used and 0 otherwise. DV1 equals 1 if high yield varieties of rice were planted, and DV2 equals 1 if mixed varieties were planted; the omitted category represents that traditional varieties were planted. DSS equals 1 if it was a wet season. Since our focus is on the efficiency estimates, regression results are not presented here (see
The LASSO estimates 69.6% of farms (119 out of 171 farms) as efficient\textsuperscript{27} The distribution of the inefficiencies are reported in Figure \ref{fig:inefficiencies}. In Figure \ref{fig:inefficiencies}, the blue histogram represents the distribution of the inefficiencies from LSDV and the orange one represents that from LASSO where the 69.6% of mass is concentrated at 0. We performed R&S procedure with error rate of 0.05, and found that 67% of rice farms (115 out of 171 farms) are in the subset of the best farms. Figure \ref{fig:subset} matches the firm IDs estimated as fully efficient by LASSO (yellow) and those included in the best subset by the R&S (red). The result is showing that all of the farms in the best subset by R&S are estimated as zero inefficiency in the LASSO. This result implies the two procedures are closely related and can be alternately used depending on the goals of research. The similarity between the results may be understood by noting that the two methods are both selecting subsets of the best firms after accounting for the impacts

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{inefficiencies.png}
\caption{Distribution of the rice farm inefficiencies}
\end{figure}

\textsuperscript{26}As in the simulations, we set $\gamma_\beta = 1$ and $\gamma_u = 2$. The model is estimated after standardizing the input variables.

\textsuperscript{27}The production function is exactly the same as Horrace and Schmidt (2000), and the LASSO selects the full model.
of the individual inefficiency estimates on the whole model, which is similar to an F-test procedure. The only difference is that the LASSO is based on the BIC criterion, whereas R&S is based on a pre-specified error rate and the multivariate confidence intervals it implies. The multivariate confidence intervals are based on all $N(N - 1)$ differences, $\tilde{\alpha}_i - \tilde{\alpha}_{i\neq j}$. If we use a different BIC criterion or a different error rate, we would get a different set of results.\textsuperscript{28}

![Figure 3.5: Firm IDs estimated as fully efficient by LASSO or included in the best subset by R&S](image)

Figure 3.5: Firm IDs estimated as fully efficient by LASSO or included in the best subset by R&S

### 3.7 Conclusion

We have shown the proposed adaptive LASSO estimator has the oracle property under regularity conditions. Moreover, the finite sample simulations demonstrate that the estimator outperforms LSDV in many aspects. The empirical application shows that our methodology and the R&S procedure produce similar results. Consequently, there may be scope for using the LASSO whenever R&S is prescribed. For example, Horrace (2005b) uses R&S to determine a subset of industries with the largest wage gap. Using Oaxaca-Blinder decomposition, the

\textsuperscript{28}In this application, the absolute values of the individual inefficiencies from the preliminary LSDV estimation were small whereas the model fit (measured by the mean squared error of the model) was bad, which leads the LASSO to keep a small number of large inefficiency estimates in the model. For the same reason, R&S resulted in wide confidence intervals, wide enough for a large portion of farms to be categorized as the best.
LASSO could be applied to select these industries.

For future research, it may be interesting to develop a zero inefficiency SF model that allows the individual inefficiencies to vary over time. The standard fixed effect model can be seen as a special case of the factor error structure panel data model when the factors are constant over time. If we assume the variation in the inefficiencies is due to time-varying factors, the factor error structure model could be a good baseline model for the analysis. One way to apply the zero inefficiency concept to the factor model would be to estimate some individual factor loadings as zeros using the shrinkage technique as in Cheng et al (2016). However, this may lead to an unsatisfying result, as some firms will have zero inefficiencies all the time. We may consider a model that allows for factor loadings to be zero for some periods but nonzero for others. However, as shown in Cheng et al (2016), this would make the model and estimation procedures prohibitively complex. Perhaps, an alternative strategy could be developed.

\[\text{For studies of the factor error structure panel model, see Pesaran (2006), Bai (2009), and Ahn et al (2013).}\]
Table 3.1: Estimation accuracy for $\hat{U}(\Pi)$

<table>
<thead>
<tr>
<th>$(N, T)$</th>
<th>$\sigma_u$</th>
<th>RMSE</th>
<th>$\hat{\alpha}$ (a_0 = 1)</th>
<th>Ranking correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LASSO</td>
<td>LSDV</td>
<td>LASSO</td>
</tr>
<tr>
<td>(100,10)</td>
<td>1</td>
<td>0.2980 (0.0370)</td>
<td>0.7591 (0.1394)</td>
<td>1.005 (0.097)</td>
</tr>
<tr>
<td>(100,30)</td>
<td>1</td>
<td>0.1840 (0.0263)</td>
<td>0.4243 (0.0786)</td>
<td>0.979 (0.051)</td>
</tr>
<tr>
<td>(100,50)</td>
<td>1</td>
<td>0.1456 (0.0214)</td>
<td>0.3252 (0.0599)</td>
<td>0.977 (0.037)</td>
</tr>
<tr>
<td>(100,70)</td>
<td>1</td>
<td>0.1223 (0.0188)</td>
<td>0.2777 (0.0533)</td>
<td>0.982 (0.032)</td>
</tr>
<tr>
<td>(100,10)</td>
<td>2</td>
<td>0.3020 (0.0416)</td>
<td>0.7390 (0.1401)</td>
<td>1.041 (0.105)</td>
</tr>
<tr>
<td>(100,30)</td>
<td>2</td>
<td>0.1762 (0.0218)</td>
<td>0.4193 (0.0795)</td>
<td>0.994 (0.049)</td>
</tr>
<tr>
<td>(100,50)</td>
<td>2</td>
<td>0.1365 (0.0162)</td>
<td>0.3219 (0.0625)</td>
<td>0.992 (0.036)</td>
</tr>
<tr>
<td>(100,70)</td>
<td>2</td>
<td>0.1143 (0.0132)</td>
<td>0.2772 (0.0528)</td>
<td>0.994 (0.030)</td>
</tr>
<tr>
<td>(100,10)</td>
<td>4</td>
<td>0.2977 (0.0429)</td>
<td>0.7266 (0.1417)</td>
<td>1.062 (0.102)</td>
</tr>
<tr>
<td>(100,30)</td>
<td>4</td>
<td>0.1699 (0.0175)</td>
<td>0.4211 (0.0820)</td>
<td>1.004 (0.046)</td>
</tr>
<tr>
<td>(100,50)</td>
<td>4</td>
<td>0.1298 (0.0133)</td>
<td>0.3274 (0.0654)</td>
<td>1.000 (0.032)</td>
</tr>
<tr>
<td>(100,70)</td>
<td>4</td>
<td>0.1102 (0.0120)</td>
<td>0.2737 (0.0529)</td>
<td>0.995 (0.027)</td>
</tr>
<tr>
<td>(200,10)</td>
<td>1</td>
<td>0.2923 (0.0264)</td>
<td>0.8240 (0.1334)</td>
<td>1.010 (0.075)</td>
</tr>
<tr>
<td>(200,70)</td>
<td>1</td>
<td>0.1216 (0.0142)</td>
<td>0.3055 (0.0511)</td>
<td>0.985 (0.025)</td>
</tr>
<tr>
<td>(200,10)</td>
<td>4</td>
<td>0.2939 (0.0263)</td>
<td>0.7917 (0.1294)</td>
<td>1.060 (0.076)</td>
</tr>
<tr>
<td>(200,70)</td>
<td>4</td>
<td>0.1093 (0.0082)</td>
<td>0.3046 (0.0502)</td>
<td>0.999 (0.020)</td>
</tr>
<tr>
<td>(1000,10)</td>
<td>1</td>
<td>0.2867 (0.0133)</td>
<td>0.9762 (0.1130)</td>
<td>1.017 (0.041)</td>
</tr>
<tr>
<td>(1000,10)</td>
<td>2</td>
<td>0.2880 (0.0103)</td>
<td>0.9751 (0.1178)</td>
<td>1.050 (0.042)</td>
</tr>
<tr>
<td>(1000,10)</td>
<td>4</td>
<td>0.2871 (0.0096)</td>
<td>0.9696 (0.1209)</td>
<td>1.068 (0.039)</td>
</tr>
</tbody>
</table>

NOTE: The entries are the average values for each measure over 1,000 replications and their corresponding standard deviations in next row in parentheses. Rank correlations are computed only among the inefficiencies whose true values are nonzero. That is, $R_{corr} = corr(R(u_{0,A}), R(\hat{U}(\Pi)_A))$ where $R(\cdot)$ is a mapping from estimates to rankings. Similarly for LSDV.
Table 3.2: Selection accuracy for $\hat{U}(\Pi)$

<table>
<thead>
<tr>
<th>(N, T)</th>
<th>$\sigma_u = 1$</th>
<th>$\sigma_u = 2$</th>
<th>$\sigma_u = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Pr_{U_A}$</td>
<td>$Pr_{U_{Ac}}$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>(100,10)</td>
<td>0.6842 (0.1051)</td>
<td>0.8843 (0.0966)</td>
<td>0.4864 (0.0950)</td>
</tr>
<tr>
<td>(100,30)</td>
<td>0.7501 (0.0853)</td>
<td>0.9440 (0.0613)</td>
<td>0.4581 (0.0713)</td>
</tr>
<tr>
<td>(100,50)</td>
<td>0.7788 (0.0762)</td>
<td>0.9589 (0.0487)</td>
<td>0.4425 (0.0620)</td>
</tr>
<tr>
<td>(100,70)</td>
<td>0.8085 (0.0704)</td>
<td>0.9594 (0.0498)</td>
<td>0.4219 (0.0580)</td>
</tr>
<tr>
<td>(200,10)</td>
<td>0.6957 (0.0796)</td>
<td>0.8792 (0.0789)</td>
<td>0.4768 (0.0749)</td>
</tr>
<tr>
<td>(200,70)</td>
<td>0.8154 (0.0526)</td>
<td>0.9575 (0.0375)</td>
<td>0.4165 (0.0441)</td>
</tr>
<tr>
<td>(1000,10)</td>
<td>0.7093 (0.0461)</td>
<td>0.8735 (0.0481)</td>
<td>0.4655 (0.0452)</td>
</tr>
</tbody>
</table>

NOTE: $Pr_{U_A}$ represents the probability of yielding nonzero estimates for $U_A$ and $Pr_{U_A}$ is computed from averaging the percentage of nonzero estimates for $U_A$ in each replication; $Pr_{U_{Ac}}$ represents the probability of yielding zero estimates for $U_{Ac}$; $\hat{\delta}^*$ represents the average proportion of the firms estimated as efficient; $Max \ U_A$ represents the maximum of $U_A$ that is estimated as zero.
References


shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(1), 63-78.


Appendices
Appendix A

Stochastic Frontier Models with Network Selectivity

A.1 Mathematical Derivations

1. $h(u_{ist} | \epsilon_{st} < \delta_{st})$

$$h(u_{ist} | \epsilon_{st} < \delta_{st}) = \frac{\int_{0}^{Z_{st}} h(\epsilon_{st}, u_{ist}) d\epsilon}{\int_{0}^{Z_{st}} \int_{0}^{\infty} h(\epsilon_{st}, u_{ist}) du d\epsilon} = \frac{1}{1 - e^{-\delta_{st}}} \int_{0}^{Z_{st}} h(\epsilon_{st}, u_{ist}) d\epsilon$$  \hfill (A.1)

Because $$\int_{0}^{Z_{st}} h(\epsilon_{st}, u_{ist}) d\epsilon = \frac{1}{\sigma_{u}} [e^{-\frac{u_{ist}}{\sigma_{u}}} - \kappa_{st} e^{-\frac{u_{ist}}{\sigma_{u}} - \delta_{st}}]$$ \text{ where } \kappa_{st} = 1 + \alpha \delta_{st},  \hfill (A.2)

$$h(u_{ist} | \epsilon_{st} < \delta_{st}) = \frac{1}{\sigma_{u} P_{st}} [e^{-\frac{u_{ist}}{\sigma_{u}}} - \kappa_{st} e^{-\frac{u_{ist}}{\sigma_{u}} - \delta_{st}}]$$ \text{ where } P_{st} = 1 - e^{-\delta_{st}} \hfill (A.3)$$
2. $E(u_{ist}|\epsilon_{st} < \delta_{st})$

\[
E(u_{ist}|\epsilon_{st} < \delta_{st}) = \int_0^\infty u_{ist} h(u_{ist}|\epsilon_{st} < \delta_{st}) du = \frac{1}{\sigma_u P_{st}} \int_0^\infty u_{ist} \left[ e^{-\frac{u_{ist}}{\sigma_u}} - \kappa_{st} e^{-\frac{u_{ist}}{\sigma_u} - \delta_{st}} \right] du
\]

\[
= \frac{\sigma_u}{P_{st}} [1 - \frac{1}{\kappa_{st}} e^{-\delta_{st}}]
\]

(A.4)

3. $h(\mu_{ist}|\epsilon_{st} < \delta_{st})$ where $\mu_{ist} = \nu_{ist} - u_{ist}$

$\nu_{ist}$ is independent of $u_{ist}$ and $\epsilon_{st}$, so $h(u_{ist}, \nu_{ist}|\epsilon_{st} < \delta_{st}) = h(u_{ist}|\epsilon_{st} < \delta_{st}) \frac{1}{\sigma_v} \phi\left(\frac{\nu_{ist}}{\sigma_v}\right)$

where $\phi(\cdot)$ is standard normal pdf, then the $h(\mu_{ist}|\epsilon_{st} < \delta_{st})$ will be given by,

(A.5)

\[
h(\mu_{ist}|\epsilon_{st} < \delta_{st}) = \int_0^\infty h(u_{ist}|\epsilon_{st} < \delta_{st}) \frac{1}{\sigma_v} \phi\left(\frac{u_{ist} + \mu_{ist}}{\sigma_v}\right) du
\]

\[
= \frac{1}{\sqrt{2\pi} \sigma_u \sigma_v P_{st}} \int_0^\infty e^{-\frac{u_{ist}}{\sigma_u} - \frac{u_{ist} + \mu_{ist}}{\sigma_u} - \frac{\sigma_v}{\sigma_u} \left(\mu_{ist} + u_{ist}\right)^2}{2\sigma_v^2} du
\]

\[
= \frac{1}{P_{st}} \left[ \frac{1}{\sqrt{2\pi} \sigma_u \sigma_v} \int_0^\infty e^{-\frac{u_{ist}}{\sigma_u} - \frac{u_{ist} + \mu_{ist}^2}{2\sigma_v^2}} du
\]

\[
- \frac{1}{\sqrt{2\pi} \sigma_u \kappa_{st} \sigma_v} \int_0^\infty e^{-\frac{\kappa_{st} u_{ist} \delta_{st} - \mu_{ist}^2}{2\sigma_v^2}} du \right]
\]

Using the result from Kumbhakar and Lovell (2000, p80),

\[
= \frac{1}{\sigma_u P_{st}} \left[ \Phi\left( -\frac{\mu_{ist}}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) e^{\frac{\mu_{ist}^2}{2\sigma_u^2} + \frac{\kappa_{st} \sigma_v^2}{2\sigma_u^2}} - \kappa_{st} \Phi\left( -\frac{\mu_{ist}}{\sigma_v} - \frac{\kappa_{st} \sigma_v}{\sigma_u} \right) e^{\frac{\kappa_{st} \mu_{ist} \sigma_v}{\sigma_u} + \frac{\kappa_{st} \sigma_v^2}{2\sigma_u^2}} \right]
\]

(A.6)
4. \( E(u^*_{\text{ist}}|\mu_{\text{ist}}, \epsilon^*_{\text{ist}} < \delta^*_{\text{ist}}) \)

\[
E(u^*_{\text{ist}}|\mu_{\text{ist}}, \epsilon^*_{\text{ist}} < \delta^*_{\text{ist}}) = \int_0^\infty u \frac{h(u_{\text{ist}}, \mu_{\text{ist}}, \epsilon^*_{\text{ist}} < \delta^*_{\text{ist}})}{h(\mu_{\text{ist}}|\epsilon^*_{\text{ist}} < \delta^*_{\text{ist}})} \, du = \int_0^\infty u \frac{h(u_{\text{ist}}, \mu_{\text{ist}}|\epsilon^*_{\text{ist}} < \delta^*_{\text{ist}})}{h(\mu_{\text{ist}}|\epsilon^*_{\text{ist}} < \delta^*_{\text{ist}})} \, du
\]

\[
= \frac{1}{\sqrt{2\pi}\sigma_v} \int_0^\infty ue^{-\frac{\mu^*_{\text{ist}} + \sigma_v^2}{2\sigma_u^2}} \, du - \frac{\kappa^*_{\text{ist}}}{\sigma_v} \int_0^\infty ue^{-\frac{\mu^*_{\text{ist}} + \sigma_v^2}{2\sigma_u^2}} \, du
\]

Let (1) = \( A_{\text{ist}}(\mu_{\text{ist}}|\delta^*_{\text{ist}}, \sigma_u, \sigma_v, \alpha) \) and (2) = \( B_{\text{ist}}(\mu_{\text{ist}}|\delta^*_{\text{ist}}, \sigma_u, \sigma_v, \alpha) \), then,

\[
= \frac{A_{\text{ist}}}{A_{\text{ist}} - B_{\text{ist}} \sqrt{2\pi}\sigma_v A_{\text{ist}}} \int_0^\infty ue^{-\frac{\mu^*_{\text{ist}} + \sigma_v^2}{2\sigma_u^2}} \, du - \frac{B_{\text{ist}}}{B_{\text{ist}} - B_{\text{ist}} \sqrt{2\pi}\sigma_v B_{\text{ist}}} \int_0^\infty ue^{-\frac{\mu^*_{\text{ist}} + \sigma_v^2}{2\sigma_u^2}} \, du
\]

Using the result from Kumbhakar and Lovell (2000, p82), this reduces to

\[
= \frac{A_{\text{ist}}}{A_{\text{ist}} - B_{\text{ist}}} \left( \Lambda^*_{\text{ist}} + \sigma_v \left[ \frac{\phi(-\Lambda^*_{\text{ist}}/\sigma_v)}{\Phi(\Lambda^*_{\text{ist}}/\sigma_v)} \right] \right) - \frac{B_{\text{ist}}}{B_{\text{ist}} - B_{\text{ist}}} \left( \Lambda^*_{\text{ist}} + \sigma_v \left[ \frac{\phi(-\Lambda^*_{\text{ist}}/\sigma_v)}{\Phi(\Lambda^*_{\text{ist}}/\sigma_v)} \right] \right)
\]

where \( \Lambda_{\text{ist}} = -\mu_{\text{ist}} - \frac{\sigma_v^2}{\sigma_u} \) and \( \Lambda^*_{\text{ist}} = -\mu_{\text{ist}} - \frac{\kappa^*_{\text{ist}} \sigma_v^2}{\sigma_u} \)

\[\text{(A.7)}\]

A.2 Three-step Estimation Approach

Three-step estimation proceeds as follows.

**Step 1:** Compute \( \delta^*_{\text{ist}} = F^{-1}(\bar{F}_{\text{st}}) \) as in the two-step approach.

**Step 2:** Do a within transformation on the production function equation \(^{1}\text{(1.2)}\) to eliminate the selection bias, and then run Conditional Maximum Likelihood estimation due to Lee

\(^{1}\text{Equation (1.2) can be rewritten as } Y_{\text{st}} = \lambda W_{\text{st}} Y_{\text{st}} + X_{1,\text{st}} \beta_1 + X_{2,\text{st}} \beta_2 1_{n_i} - E(u_{\text{ist}}|\epsilon^*_{\text{ist}} < \delta^*_{\text{ist}}) + \mu^*_{\text{ist}}, \text{ where } \mu^*_{\text{ist}} = \nu_{\text{st}} - u^*_{\text{ist}} + E(u_{\text{ist}}|\epsilon^*_{\text{ist}} < \delta^*_{\text{ist}}). \text{ Then, the within transformation on the equation leads to } QY_{\text{st}} = Q\lambda W_{\text{st}} Y_{\text{st}} + QX_{1,\text{st}} \beta_1 + Q\mu^*_{\text{ist}}, \text{ where } Q \text{ is the within transformation matrix. Now the error term } Q\mu^*_{\text{rs}} \text{ has zero mean by construction.} \)
(2007) or 2SLS to get \( \hat{\lambda} \) and \( \hat{\beta}_1 \).

Using the efficient instruments of IV = \((W_{st}X, W_{st}^2X, \ldots)\) for the right-hand side endogenous variables, \( W_{st}Y_{st} \), is a standard method for 2SLS estimation of \( \hat{\lambda} \) in spatial production literature. See Anselin (1988) or Kelejian and Prucha (1999). However, the applicability of this Fixed effect-2SLS methodology heavily depends on the topology of the weight matrix, and this is closely related to the identification problem in network models. Since Bramoulle et al (2009) and Cohen et al (2012) extensively studied sufficient identification conditions for various network models, we won’t discuss them again here, rather we will show how network topology affects the applicability of the Fixed effect-2SLS. Let’s assume the network of group \( s \) in time \( t \) is given by

\[
W_{st}^* = \frac{1}{N_t-1}(1_{N_t}1_{N_t}' - I_{N_t})
\]

and \( N_t \) is constant over \( t \), where \( N_t \) is the number of the agents in a chosen group. This weight matrix has been widely used in the spatial autoregressive model due to its simplicity, however, if we apply a within transformation on the network, we get \( QW_{st}^* = -\frac{1}{N_t-1}Q \), then the set of IVs becomes \((QX, QW_{st}^*X, \ldots) = (QX, -\frac{1}{N_t-1}QX, \ldots)\), which implies the set of IVs is not full rank and suffers from a perfect multicollinearity problem as long as \( N_t \) doesn’t vary over \( t \), which, in turn, leads to a failure of 2SLS. For this reason, HLP introduce two heterogeneous weight matrices using exclusion restrictions, \( W_{st}^1 \) and \( W_{st}^2 \), the same- and different-type weight matrices, respectively, where \( W_{0, st}^1 = [w_{0, ij, st}] \) is an adjacency matrix with \( w_{0, ij, st} = 1 \) if the \( i^\text{th} \) and \( j^\text{th} \) workers in the group \( s \) in time \( t \) are of the same type and \( w_{0, ij, st} = 0 \) otherwise. Similarly construct \( W_{0, st}^2 \), then row-normalize \( W_{0, st}^1 \) and \( W_{0, st}^2 \) to produce \( W_{st}^1 \) and \( W_{st}^2 \) such that \( W_{st}^11_{N_t} = W_{st}^21_{N_t} = 1_{N_t} \).

In short, in order to implement 3-step estimation, \( N_t \) should vary over \( t \), or the weight matrices should be different enough from \( W_{st}^* \). If those issues are addressed, we may apply CML or 2SLS.

**Step 3:** Using the estimates, \( \hat{Z}_{ts}, \hat{\lambda} \) and \( \hat{\beta}_1 \) from above steps, compute \( \hat{r}_{ts} = Y_{ts} - \hat{\lambda}W_{ts}Y_{ts} - X_{1,ts}\hat{\beta}_1 \), then the consistent estimate for the \( \mu_{ts} \) conditional on the \( x_{2,ts} \) and \( \beta_2 \) is given by
\[ \hat{\mu}_{ts} = \hat{r} - \chi_{2,ts} \beta_2. \]

Plug the \( \hat{\mu}_{ts} \) into the LL in (1.7) without the term \( \sum_t \ln |l_{N_t} - \lambda W_{ts}| \) because we already removed endogeneity due to the spatial autoregressive component in the right side of the production equation from the second step, and run the Pseudo-ML. Here we get \( \hat{\beta}_2, \hat{\sigma}_u, \hat{\sigma}_v, \) and \( \hat{\alpha} \). As in the two step method, we use Murphy and Topel (2002) to correct the standard errors.

### A.3 The partial and the second derivatives of the log-likelihood and adjustment of the standard error

\[
\begin{align*}
\ln L = & \sum_t \sum_{i \in s} \ln h(\epsilon_{st} < \delta_{st}) h(\mu_{ist}|\epsilon_{st} < \delta_{st}) = \sum_t N_t \ln \frac{\eta}{\sigma_v} + \sum_t \ln |l_{N_t} - \lambda W_{st}| \\
& + \sum_t \sum_{i \in s} \ln \left[ \Phi \left( -\frac{\mu_{ist}}{\sigma_v} - \eta \right) e^{\frac{\mu_{ist}^2}{2\sigma_v^2} + \frac{1}{2}\eta^2} - \frac{\kappa_{ist}}{e^{\delta_{st}}} \Phi \left( -\frac{\mu_{ist}}{\sigma_v} - \kappa_{ist}\eta \right) e^{\frac{\kappa_{ist}\mu_{ist}^2}{2\sigma_v^2} + \frac{\kappa_{ist}^2\eta^2}{2}} \right]
\end{align*}
\]

#### 1. Notation

For the simplicity, we use below notation,

\[
\omega_{ist} = -\frac{\mu_{ist}}{\sigma_v} - \eta, \quad \omega_{kst} = -\frac{\mu_{ist}}{\sigma_v} - \kappa_{ist}\eta \tag{A.8}
\]

\[
\phi_{ist} = \phi(\omega_{ist}), \quad \phi_{kst} = \phi(\omega_{kst}), \quad \Phi_{ist} = \Phi(\omega_{ist}), \quad \Phi^{kst} = \Phi(\omega_{kst}) \tag{A.9}
\]

\[
\tau_{ist} = e^{\frac{\mu_{ist}^2}{2\sigma_v^2} + \frac{1}{2}\eta^2}, \quad \tau^{kst} = e^{\frac{\kappa_{ist}\mu_{ist}^2}{2\sigma_v^2} + \frac{\kappa_{ist}^2\eta^2}{2}} \tag{A.10}
\]

\[
\psi_{ist} = \Phi_{ist} \tau_{ist} - \frac{\kappa_{ist}}{e^{\delta_{st}}} \Phi^{kst} \tau^{kst} \tag{A.11}
\]
2. The first derivatives

\[
\frac{\partial \ln L}{\partial \lambda} = - \sum_t \text{tr}[(I_n - \lambda W_t)^{-1} W_t] + \sum_t \sum_i \frac{\left(\sum_{j \neq i \in s} W_{jt} y_{ist}\right)}{\sigma_v} \xi_{ist} \Psi_{ist}
\]

where \(\xi_{ist} = \tau_{ist}(\phi_{ist} - \eta \Phi_{ist}) - \kappa_{ist}^\kappa \tau_{ist}^\kappa(\phi_{ist}^\kappa - \kappa_{ist}^\kappa \eta \Phi_{ist}^\kappa)\)

\[
\frac{\partial \ln L}{\partial \eta} = \sum_t \frac{N_t}{\eta} - \sum_t \sum_i \frac{\xi_{ist}^\eta}{\Psi_{ist}}
\]

where \(\xi_{ist}^\eta = \tau_{ist}(\omega_{ist} \Phi_{ist} + \phi_{ist}) - \kappa_{ist}^2 \tau_{ist}^\kappa(\omega_{ist}^\kappa \Phi_{ist}^\kappa + \phi_{ist}^\kappa)\)

\[
\frac{\partial \ln L}{\partial \sigma_v} = - \sum_t \frac{N_t}{\sigma_v} + \sum_t \sum_i \mu_{ist} \frac{\xi_{ist}}{\sigma_v^2} \Psi_{ist}
\]

\[
\frac{\partial \ln L}{\partial \beta} = \sum_t \sum_i \frac{\tau_{ist}}{\sigma_v} \frac{\xi_{ist}}{\Psi_{ist}}
\]

\[
\frac{\partial \ln L}{\partial \alpha} = - \sum_t \sum_i \frac{\tau_{ist}}{\sigma_v} \frac{\xi_{ist}}{\Psi_{ist}}
\]

where \(\xi_{ist} = \Phi_{ist}^\kappa - \kappa_{ist} \eta(\omega_{ist}^\kappa \Phi_{ist}^\kappa + \phi_{ist}^\kappa)\)

3. The second derivatives

\[
\frac{\partial^2 \ln L}{\partial \lambda^2} = - \sum_t \text{tr}[(I_n - \lambda W_t)^{-1} W_t]^2 - \sum_t \sum_i \left(\frac{\left(\sum_{j \neq i \in s} W_{jt} y_{ist}\right)}{\sigma_v}\right)^2
\]

\[
\begin{aligned}
&\left\{ \left(\frac{\xi_{ist}}{\Psi_{ist}}\right)^2 - \frac{\tau_{ist}(\frac{\mu_{ist}}{\sigma_v} - \eta) \phi_{ist} + \eta^2 \Phi_{ist}}{\Psi_{ist}} - \frac{\kappa_{ist}^2 \tau_{ist}^\kappa(\frac{\mu_{ist}}{\sigma_v} - \kappa_{ist} \eta) \phi_{ist}^\kappa + \kappa_{ist}^2 \eta^2 \Phi_{ist}^\kappa}{\Psi_{ist}} \right\} \\
&\left(\frac{\xi_{ist}}{\Psi_{ist}}\right) \left(\frac{\tau_{ist}}{\sigma_v} \left(\frac{\mu_{ist}^2}{\sigma_v} - \eta^2 \Phi_{ist}^2\right) - \kappa_{ist}^2 \tau_{ist}^\kappa \phi_{ist}^\kappa - \kappa_{ist}^2 \eta^2 \Phi_{ist}^\kappa\right)
\end{aligned}
\]

(A17a)

\[
\frac{\partial^2 \ln L}{\partial \lambda_1 \partial \lambda_2} = - \sum_t \text{tr}[(I_n - \lambda_1 W_t^1 - \lambda_2 W_t^2)^{-1} W_t^1 (I_n - \lambda_1 W_t^1 - \lambda_2 W_t^2)^{-1} W_t^2]
\]

\[
- \sum_t \sum_i \left(\frac{\left(\sum_{j \neq i \in s} W_{jt} y_{ist}\right)}{\sigma_v^2} \sum_{j \neq i \in s} W_{jt}^2 y_{ist} \right) \times Q_{ist}
\]

(A17b)
\[
\frac{\partial^2 \ln L}{\partial \eta^2} = -\sum_t N_t - \sum_t \sum_{i \in S} \left\{ \left( \frac{\xi_{ist}^\eta}{\psi_{ist}} \right)^2 \right. \\
- \tau_{ist} \left( \omega_{ist}^2 + (\omega_{ist}^2 + 1) \phi_{ist} \right) - \frac{\kappa_{ist}^3}{\sigma_v} \tau_{ist}^\eta \left( \omega_{ist}^2 + (\omega_{ist}^2 + 1) \phi_{ist} \right) \left( \frac{\xi_{ist}^\eta}{\psi_{ist}} \right)^2 \left\} \right.
\]

(A.18)

\[
\frac{\partial^2 \ln L}{\partial \sigma_v^2} = -\sum_t N_t - \sum_t \sum_{i \in S} \left( \frac{\mu_{ist}}{\sigma_v^2} \right)^2 \times Q_{ist} 
\]

(A.19)

\[
\frac{\partial^2 \ln L}{\partial \beta^2} = -\frac{1}{\sigma_v^2} \sum_t \sum_{i \in S} \chi_{ist} \chi_{ist}' \times Q_{ist} 
\]

(A.20)

\[
\frac{\partial^2 \ln L}{\partial \alpha^2} = -\sum_t \sum_{i \in S} \left\{ \left( \frac{\xi_{ist}^\alpha}{e^{\delta_{ist}}} \right) \frac{\xi_{ist}}{\psi_{ist}} \right. \\
- \frac{\tau_{ist}^\alpha \omega_{ist}^2}{\psi_{ist}} \left( \omega_{ist}^2 + (\omega_{ist}^2 + 1) \phi_{ist} \right) + 2\phi_{ist}^\alpha + \phi_{ist}^\alpha \left( \omega_{ist}^2 + \kappa_{ist}^\alpha \right) \left( \frac{\xi_{ist}}{\psi_{ist}} \right)^2 \left\} \right.
\]

(A.21)

\[
\frac{\partial^2 \ln L}{\partial \lambda \partial \eta} = \sum_t \sum_{i \in S} \left( \sum_{j \neq i \in S} W_{tj} y_{ist} \right) \sigma_v \\
\left\{ \frac{\xi_{ist}^\eta}{\psi_{ist}^2} + \tau_{ist} \left( \omega_{ist} \phi_{ist} - \phi_{ist} + \eta \phi_{ist} \right) - \frac{\kappa_{ist}^2}{\sigma_v} \tau_{ist}^\eta \left( \omega_{ist}^2 + (\omega_{ist}^2 + 1) \phi_{ist} \right) \left( \frac{\xi_{ist}}{\psi_{ist}} \right)^2 \left\} \right.
\]

(A.22)

\[
\frac{\partial^2 \ln L}{\partial \lambda \partial \sigma_v} = -\sum_t \sum_{i \in S} \left( \sum_{j \neq i \in S} W_{tj} y_{ist} \right) \sigma_v \left\{ \frac{\xi_{ist}}{\psi_{ist}} + \frac{\mu_{ist}}{\sigma_v} \times Q_{ist} \right\} 
\]

(A.23)

\[
\frac{\partial^2 \ln L}{\partial \lambda \partial \beta^2} = -\sum_t \sum_{i \in S} \left( \sum_{j \neq i \in S} W_{tj} y_{ist} \right) \chi_{ist}' \times Q_{ist} 
\]

(A.24)
\[ \frac{\partial^2 \ln L}{\partial \lambda \partial \alpha} = \sum_t \sum_{i \in s} \left( \sum_{j \neq i \in s} W_{ij} y_{ist} \right) \frac{\tau_{ist}^\kappa \delta_{st}}{\sigma_v} \psi_{ist}^2 \left\{ \frac{\xi_{ist}^\alpha \psi_{ist}^\alpha}{\psi_{ist}^2} - \frac{\phi_{ist}^\kappa (1 + \kappa_{st} \eta^2) + \Phi_{ist}^\kappa (\kappa_{st}^2 \omega_{ist} \eta^2 - 2 \kappa_{st} \eta)}{\psi_{ist}} \right\} \] (A.25)

\[ \frac{\partial^2 \ln L}{\partial \eta \partial \sigma_v} = \sum_t \sum_{i \in s} \frac{\mu_{ist}}{\sigma_v^2} \times R_{ist} \] (A.26)

\[ \frac{\partial^2 \ln L}{\partial \eta \partial \beta'} = \sum_t \sum_{i \in s} \frac{\chi_{ist}'}{\sigma_v} \times R_{ist} \] (A.27)

\[ \frac{\partial^2 \ln L}{\partial \eta \partial \alpha} = -\sum_t \sum_{i \in s} \frac{\tau_{ist}^\kappa \delta_{st}}{\sigma_v^2} \left\{ \frac{\xi_{ist}^\alpha \psi_{ist}^\alpha}{\psi_{ist}^2} - \kappa_{ist} \omega_{ist} \xi_{ist}^\alpha + 2 \phi_{ist}^\kappa \omega_{ist} (\omega_{ist}^\alpha - \kappa_{st} \eta) \right\} \] (A.28)

\[ \frac{\partial^2 \ln L}{\partial \sigma_v \partial \beta} = -\sum_t \sum_{i \in s} \frac{\chi_{ist}}{\sigma_v} \left\{ \frac{\xi_{ist}}{\psi_{ist}} + \frac{\mu_{ist}}{\sigma_v} \times Q_{ist} \right\} \] (A.29)

\[ \frac{\partial^2 \ln L}{\partial \sigma_v \partial \alpha} = \sum_t \sum_{i \in s} \frac{\mu_{ist}}{\sigma_v^2} \frac{\tau_{ist}^\kappa \delta_{st}}{\sigma_v^2} \times S_{ist} \] (A.30)

\[ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} = \sum_t \sum_{i \in s} \frac{\chi_{ist}}{\sigma_v} \frac{\tau_{ist}^\kappa \delta_{st}}{\sigma_v} \times S_{ist} \] (A.31)

4. Adjustment of the Hessian for multi-step estimation

Following Murphy and Topel (2002), the adjusted standard error is given by,

\[ \Sigma = R_2^{-1} + R_2^{-1} [R_1^3 R_1^{-1} R_3 - R_1^3 R_1^{-1} R_3 - R_1^3 R_1^{-1} R_4] R_2^{-1} \] (A.32)
\[
R_1 = -E \frac{\partial^2 \ln L_1}{\partial \theta_1 \partial \theta_1'} \\
R_2 = -E \frac{\partial^2 \ln L_2}{\partial \theta_2 \partial \theta_2'} \\
R_3 = -E \frac{\partial^2 \ln L_2}{\partial \theta_1 \partial \theta_2'} \\
R_4 = E \frac{\partial \ln L_1}{\partial \theta_1} \left( \frac{\partial \ln L_2}{\partial \theta_2} \right)
\]

(A.33)

where \(\ln L_1\), and \(\theta_1\) are the log-likelihood and the parameters from the first step, and the \(\ln L_2\), and \(\theta_2\) are the log-likelihood and the parameters from the second step. \(R_1\) and \(R_2\) can be obtained from the Hessian matrix of the conditional logit model and the second derivatives in subsequent derivations. As the log-likelihood for the first step is given by,

\[
\ln L_1 = \sum_t \left( Z_{st}\gamma - \ln \sum_{j \in At} e^{Z_{jt}\gamma} \right) + C
\]

(A.34)

where \(\gamma\) is the parameter in the first step, \(Z_{st}\) is explanatory variable for line-up \(s\) in time \(t\), \(At\) is the set of possible line-ups in \(t\), and \(C\) is the constant which does not depend on the \(\gamma\), the first derivative of the first selection model is given by,

\[
\frac{\partial \ln L_1}{\partial \gamma} = \sum_t \left( Z_{st} - \sum_{j \in At} P_{jt}Z_{jt} \right)
\]

(A.35)

where \(P_{jt}\) is given by \(P_{jt} = \frac{e^{Z_{jt}\gamma}}{\sum_{k \in At} e^{Z_{kt}\gamma}}\). Then \(R_4\) can be computed from (A.35) and the earlier first derivatives. For \(R_3\), using \(\frac{\partial \delta_{st}}{\partial \gamma} = \frac{\partial F^{-1}(P_s)}{\partial \gamma} = \frac{1}{e^{-\delta_{st}}} \Pi_{st} \) where \(\Pi_{st} = P_{ts}(m_{ts} - \sum_{j \in At} P_{ty}m_{ty})\), we have,

\[
\frac{\partial \ln L_2}{\partial \gamma} = -\sum_t \sum_{i \in s} \Pi_{st} \frac{\tau^{\kappa_{st}}(\alpha^{\epsilon_{ist}} - \kappa_{st} \Phi^{\kappa_{ist}})}{\psi_{ist}}
\]

(A.36)
Then, the cross derivatives are given by,

\[
\frac{\partial^2 \ln L_2}{\partial \gamma \partial \lambda} = \sum_t \sum_{i \in s} \Pi_{ts} \frac{\left( \sum_{j \neq i \in s} W_{ij} y_{tj} \right) T_{ist}^{\kappa_{ist}}}{\sigma_v} \\
\left\{ \left( \frac{\alpha \xi_{ist}^{\alpha} - \kappa_{ist} \Phi_{ist}^{\kappa_{ist}}}{\psi_{tsi}^2} \right) \xi_{ist}^{\alpha} - \frac{\alpha \kappa_{ist} \eta \xi_{ist}^{\alpha}}{\psi_{tsi}^2} - \kappa_{ist} \eta \left( \kappa_{ist} \Phi_{ist}^{\kappa_{ist}} - \alpha \Phi_{ist}^{\kappa_{ist}} \right) + \kappa_{ist} \phi_{ist}^{\kappa_{ist}} - \alpha \phi_{ist}^{\kappa_{ist}} \right\}
\]

(A.37)

\[
\frac{\partial^2 \ln L_2}{\partial \gamma \partial \eta} = -\sum_t \sum_{i \in s} \Pi_{ts} T_{ist}^{\kappa_{ist}} \left\{ \left( \frac{\alpha \xi_{ist}^{\alpha} - \kappa_{ist} \Phi_{ist}^{\kappa_{ist}}}{\psi_{tsi}^2} \right) \xi_{ist}^{\alpha} - \kappa_{ist} \alpha \omega_{ist} \xi_{ist}^{\alpha} - \Phi_{ist}^{\kappa_{ist}} \omega_{ist} \kappa_{ist} - \kappa_{ist} \phi_{ist}^{\kappa_{ist}} + \alpha \left( 2 \phi_{ist}^{\kappa_{ist}} + \omega_{ist} \Phi_{ist}^{\kappa_{ist}} - \eta \kappa_{ist} \Phi_{ist}^{\kappa_{ist}} \right) \right\}
\]

(A.38)

\[
\frac{\partial^2 \ln L_2}{\partial \gamma \partial \sigma_v} = \sum_t \sum_{i \in s} \Pi_{ts} T_{ist}^{\kappa_{ist}} \frac{\mu_{ist}}{\sigma_v^2} T_{ist}
\]

(A.39)

\[
\frac{\partial^2 \ln L_2}{\partial \gamma \partial \beta'} = \sum_t \sum_{i \in s} \Pi_{ts} T_{ist}^{\kappa_{ist}} \frac{\chi'_{ist}}{\sigma_v} T_{ist}
\]

(A.40)

\[
\frac{\partial^2 \ln L_2}{\partial \gamma \partial \alpha} = -\sum_t \sum_{i \in s} \Pi_{ts} T_{ist}^{\kappa_{ist}} \left\{ \frac{T_{ist}^{\kappa_{ist}} \delta_{ist} \left( \frac{\alpha \xi_{ist}^{\alpha} - \kappa_{ist} \Phi_{ist}^{\kappa_{ist}}}{\psi_{tsi}^2} \right) \xi_{ist}^{\alpha}}{\delta \xi_{ist}^{\alpha}} \right\}
\]

\[
\left\{ \xi_{ist}^{\alpha} \left( 1 - \alpha \omega_{ist}^{\kappa_{ist}} \delta_{ist} \eta \right) + \delta_{ist} \eta \Phi_{ist}^{\kappa_{ist}} \left( \omega_{ist}^{\kappa_{ist}} \kappa_{ist} - \alpha \omega_{ist}^{\kappa_{ist}} - \alpha \eta \kappa_{ist} \right) + \delta_{ist} \eta \phi_{ist}^{\kappa_{ist}} \left( \kappa_{ist} - 2 \alpha \right) - \delta \Phi_{ist}^{\kappa_{ist}} \right\}
\]

(A.41)
Appendix B

Network Competition and Team Chemistry in the NBA

B.1 GMM for the heterogeneous network interaction model

Here, we provide a sketch of the GMM method for the $R = 2$ case. With slightly modified notation for the simplicity of the case, the model is

$$ Y = \sum_{r=1}^{4} \lambda_r G_r Y + X\beta + A + U^* \quad (B.1) $$

where $G_1 = \begin{bmatrix} W_{11} & 0 \\ 0 & 0 \end{bmatrix}$, $G_2 = \begin{bmatrix} 0 & W_{22} \\ 0 & 0 \end{bmatrix}$, $G_3 = \begin{bmatrix} 0 & 0 \\ W_{21} & 0 \end{bmatrix}$, $G_4 = \begin{bmatrix} 0 & 0 \\ 0 & W_{22} \end{bmatrix}$. To eliminate the network specific fixed effects, we premultiply by $J_Q$,

$$ J_Q Y = \sum_{r=1}^{4} \lambda_r J_Q G_r Y + J_Q X\beta + J_Q U \quad (B.2) $$

Kelejian and Prucha (1999, 2001) introduce a method of moments (MOM) estimator for the SAR disturbance model, and Lee (2001a, 2002b, 2007b) develop it within a general GMM estimation framework and propose a GMM method which explores both IV (linear) as well
as quadratic moment functions\(^1\) They derive a best GMM estimator (BGMME) and show it to have the same limiting distribution as the MLE or QML estimator. Lee and Liu (2010) extend the GMM methodology to the higher-order SAR model.

We assume that each element in \( U \) is \( i.i.d. \mathcal{N}(0, \sigma^2) \) to simplify the estimation\(^2\). Following Lee’s methodology, the \( i.i.d. \mathcal{N}(0, \sigma^2) \) disturbances imply two sets of moment conditions: linear and quadratic

\[
g(\theta) = \begin{pmatrix}
Z'J_QU(\theta) \\
U'(\theta)J_QF_{1J_QU(\theta)} \\
\vdots \\
U'(\theta)J_QF_{rJ_QU(\theta)}
\end{pmatrix}
\]

\( J_QZ \) is a matrix instrument for the linear moments, and \( F_rJ_QU(\theta) \) is for the quadratic moments, where \( F_r \) is an \( n \times n \) matrix such that \( \text{tr}(F_iJ_Q) = 0 \), and \( U(\theta) = Y - \sum_{i=1}^{4} \lambda_i G_i Y + X\beta \) with \( \theta = (\lambda', \beta')' \). At \( \theta_0 \), \( E(g(\theta_0)) = 0 \) as \( E(Z'J_QU) = 0 \) and \( E(U'J_QF_{ij}U) = \sigma^2 \text{tr}(F_iJ_Q) = 0 \) for \( i = 1...m \). We can find the intuition for the use of quadratic moments in addition to the linear moments from Lee (2001a, 2001b). In (B.2), each \( J_QG_iY \) can be expressed as \( J_QG_i(I_n - \sum_{i=1}^{4} \lambda_i G_i)^{-1}X\beta + J_QG_i(I_n - \sum_{i=1}^{4} \lambda_i G_i)^{-1}U \), which has the deterministic part and the stochastic part. Then, when we construct moment conditions for the endogenous variables, \( J_QG_iY \), we can use the function of the exogenous variables, \( X \), for the approximation of the deterministic part (linear moments) while using the quadratic moments for the stochastic part as long as the \( F_iJ_QU \) and the stochastic part are correlated.

The GMM estimator \( \hat{\theta} \) follow from \( \arg \min_{\theta} g'(\theta)Hg(\theta) \), where \( H \) is a distance matrix (or

---

\(^1\)See Lee and Lin (2010), Lee et al (2010B), Lee and Yu (2014) for additional details.

\(^2\)The normality assumption is for the simpler forms of the Best IVs and the corresponding estimators’ asymptotic distribution later. We can relax the normality and homoskedasticity assumptions in the error terms as long as there is no correlation between the error terms to apply the Lee’s methodology. See Lee et al (2010b) for non-normality assumption, and Lee and Lin (2010) for the heteroskedastic specification.

\(^3\)If there is a time dimension of \( 1...T \), \( Z = (Z'_1, ..., Z'_T)' \), \( U = (U'_1, ..., U'_T)' \), \( J_Q,T = I_T \otimes J_Q \), and \( F_{i,T} = I_T \otimes F_i \).
weighting matrix) for the system of equations. In practice, we need to select specific \( Z \), \( F_i \)'s, and \( H \) to implement the GMM. The choice of \( H \) would be rather easy because, following Hansen’s setting and approach (1982), the optimal \( H \) will be the inverse of the variance matrix of the moment conditions with the chosen \( Z \), \( F_i \)'s.\(^4\) For \( Z \), \( F_i \)'s, we may consider of the Lee(2007b)'s best \( Z \) and \( F_i \)'s which has the least variance in the asymptotic distribution of \( \hat{\theta} \). For our model, these are

\[
\begin{align*}
\square & \quad J_Q Z = \left[ J_Q X, J_Q C_1 X \hat{\beta}, \ldots, J_Q C_4 X \hat{\beta} \right] \text{ where } C_i = G_i (I_n - \sum_{i=1}^{4} \hat{\lambda}_i G_i)^{-1} \text{ where } \hat{\lambda}_i \text{ and } \hat{\beta} \text{ are the preliminary estimates for } \lambda \text{ and } \beta. \\
\square & \quad F = (F_i) \text{ for } i = 1\ldots 4 \text{ where } F_i = C_i - \frac{tr(C_i J_Q)}{n-2} J_Q
\end{align*}
\]

We denote this as GMM_{Lee}. The derivation of the best \( Z \), \( F_i \)'s and the asymptotic distribution of the estimators can be found in Appendix B.4. GMM_{Lee} requires the preliminary estimates of \( \hat{\lambda}_i \)'s and \( \hat{\beta} \) as well as the estimate of the variance matrix of the moment conditions. So, alternatively, we may also think of simpler Kelejian and Prucha (2004)'s linear moments and their corresponding quadratic moments (GMM_{KP}) as below\(^5\)

\[
\begin{align*}
\square & \quad J_Q Z = \left[ J_Q X, J_Q G_1 X, J_Q G_2 X, \ldots, J_Q G_1\hat{s}_1 G_2\hat{s}_2 G_3\hat{s}_3 G_4\hat{s}_4 X \right] s_1, s_2, s_3, s_4, \geq 0 \\
\square & \quad F = (F_i) \text{ for } i = 1\ldots m \text{ where } F_i = H_i - \frac{tr(H_i J_Q)}{n-2} J_Q \text{ where } H_i = G_1\hat{s}_1 G_2\hat{s}_2 G_3\hat{s}_3 G_4\hat{s}_4, s_1, s_2, s_3, s_4, \geq 0
\end{align*}
\]

In order to reduce the number of parameters in \( g' (\theta) \hat{\Omega}^{-1} g(\theta) \) where \( \theta = (\lambda', \beta')' \), we may replace \( \beta \) with consistent estimator, \( \hat{\beta}(\lambda) = [X' J_Q X]^{-1} X' J_Q (I_n - \sum_{i=1}^{4} \lambda_i G_i) Y \). Lee (2007c) analyzes the effect of this modification on the asymptotic distribution of the estimators.

\(^4\)This should be estimated from the estimated residuals of \( U \) from an initial consistent estimate of \( \theta \).

\(^5\)The asymptotic distribution for this estimator can be established similarly as in the Appendix B.4.
B.2 Identification condition for the equation (2.11)

**Proposition B.2.1** The equation (2.11) is identified if \([J_{Q,m}X_{k,m}, J_{Q,m}Θ_{k,1,m}, \ldots, J_{Q,m}Θ_{k,n_m,m}]\) has a full rank for \(∀k = 1, \ldots, n_m\) in some \(m\) where \(Θ_{i,j,m} = G_{i,j,m}S_{m}^{-1}X_mB\).

**Proof** Denote \(S_m = I_{Nm} - \sum_{k=1}^{n_m} \sum_{l=1}^{n_m} λ_{k,l}G_{k,l,m}\). As \(S_m^{-1} = I_{Nm} + \sum_{k=1}^{n_m} \sum_{l=1}^{n_m} λ_{k,l}G_{k,l,m}S_m^{-1}\) and \(J_{Q,m}Y = J_{Q,m}S_m^{-1}X_mB + J_{Q,m}S_m^{-1}U_m^*\), the equation (2.11) can be written as

\[
J_{Q,m}Y_m = J_{Q,m}S_m^{-1}X_mB + J_{Q,m}S_m^{-1}U_m^*
= \sum_{k=1}^{n_m} \sum_{l=1}^{n_m} λ_{k,l}J_{Q,m}G_{k,l,m}S_m^{-1}X_mB + J_{Q,m}X_mB + J_{Q,m}S_m^{-1}U_m^*
\]

(B.3)

Denote \(Θ_{i,j,m} = G_{i,j,m}S_m^{-1}X_mB\), then, we can see (B.3) will be identified as long as

\[
[J_{Q,m}X_m, J_{Q,m}Θ_{1,1,m}, \ldots, J_{Q,m}Θ_{1,n_m,m}, \ldots, J_{Q,m}Θ_{n_m,1,m}, \ldots J_{Q,m}Θ_{n_m,n_m,m}]
\]

(B.4)

has a full rank. Due to the structures of the network matrices, the rank condition is equivalent to that

\[
[J_{Q,m}X_{k,m}, J_{Q,m}Θ_{k,1,m}, \ldots, J_{Q,m}Θ_{k,n_m,m}]
\]

(B.5)

has a full rank for \(∀k = 1, \ldots, n_m\) in some \(m\) where \(X_{k,m}\) is a \(N_m \times p\) matrix whose entries in the position \((\sum_{k=1}^{i-1} n_{k,m} + 1 : \sum_{k=1}^{i} n_{k,m}, 1 : p)\) is given by \(x_{i,m}\) but the rest entries are zeros so that \(\sum_{k=1}^{n_m} X_{k,m} = (x_{1,m}', \ldots, x_{n_m,m}')^\top\) where \((a : b, c : d)\) means entries from the \(a^{th}\) row to \(b^{th}\) row and from \(c^{th}\) column to \(d^{th}\) column. This condition is generally satisfied because we have multiple sets of network matrices and exogenous regressors for each group which produces enough variations to identify the coefficients in our model. ■

\(^6\)Note that \(J_{Q,m}G_{k,l,m}J_{Q,m} = J_{Q,m}G_{k,l,m} \quad ∀ k,j\), which leads that \(J_{Q,m}S_m^{-1}J_{Q,m} = J_{Q,m}S_m^{-1}\)
**Lemma B.3.1** Let the orthonormal matrix of $Q_{i,m}$ be $[P_{i,m}, l_{n,m}/\sqrt{n_{i,m}}]$. The columns in $P_{i,m}$ are eigenvectors of $Q_{i,m}$ corresponding to the eigenvalue one, such that $P_{i,m} l_{n,m} = 0$, $P_{i,m} P_{i,m} = l_{n,m}^{-1}$ and $P_{i,m} P_{i,m}' = Q_{i,m}$. Denote $J_{P,m} = \text{Diag}(P_{1,m}', ..., P_{n,m}')$ and $\tilde{G}_{i,j,m} = J_{P,m} G_{i,j,m} J_{P,m}$, then, $\ln |\tilde{S}_m| = -\ln f(\Lambda) + \ln |S_m|$ where $\tilde{S}_m = l_{N_m-n_m} - \sum_{k=1}^{n_m} \sum_{l=1}^{n_m} \lambda_{k,l} \tilde{G}_{k,l,m}$, $S_{k,m} = l_{N_m} - \sum_{k=1}^{n_m} \sum_{l=1}^{n_m} \lambda_{k,l} G_{k,l,m}$ and $f(\Lambda)$ is some function of $\Lambda$.

**Proof** Here, we show that this lemma holds for two networks case. From this, we can easily see that the Lemma holds for any number of networks case. We suppress $m$ here. Define

$$H = \begin{bmatrix} P_1, l_n/\sqrt{m} \\ 0 \\ P_2, l_n/\sqrt{m} \end{bmatrix}.$$ 

Then, we can show that $|H'(I_N - \sum_{k=1}^{2} \sum_{l=1}^{2} \lambda_{k,l} G_{k,l})H| = |H'H||I_N - \sum_{k=1}^{2} \sum_{l=1}^{2} \lambda_{k,l} G_{k,l}| = |I_N - \sum_{k=1}^{2} \sum_{l=1}^{2} \lambda_{k,l} G_{k,l}|$ as $|H'H| = 1$. Next, we show

$$H'(I_N - \sum_{k=1}^{2} \sum_{l=1}^{2} \lambda_{k,l} G_{k,l})H = \begin{bmatrix} P_1, l_n/\sqrt{m} \\ 0 \\ P_2, l_n/\sqrt{m} \end{bmatrix} \begin{bmatrix} I_{n_1} - \lambda_{1,1} W_{1,1} & -\lambda_{1,2} W_{1,2} \\ -\lambda_{2,1} W_{2,1} & I_{n_2} - \lambda_{2,2} W_{2,2} \end{bmatrix} \begin{bmatrix} P_1, l_n/\sqrt{m} \\ 0 \\ P_2, l_n/\sqrt{m} \end{bmatrix}$$

as $P_i' W_{i,j} P_j = 0$ and $l_n' W_{i,j} l_n = n_i$ for $i,j = 1,2$

but $l_n' W_{i,j} P_j$ may not be zero because $W_{i,j}$ is not necessarily symmetric.

$$= \begin{bmatrix} P_1'(I_{n_1} - \lambda_{1,1} W_{1,1}) P_1 & 0 & -\lambda_{1,2} P_1' W_{1,2} P_2 & 0 \\ l_n' \sqrt{m} (I_{n_1} - \lambda_{1,1} W_{1,1}) P_1 & 1 - \lambda_{1,1} & l_n' \sqrt{m} (-\lambda_{1,2} W_{1,2}) P_2 & -\sqrt{m} \sqrt{n_2} \lambda_{1,2} \\ -\lambda_{2,1} P_2' W_{2,1} P_1 & 0 & P_2'(I_{n_2} - \lambda_{2,2} W_{2,2}) P_2 & 0 \\ l_n' \sqrt{n_2} (-\lambda_{2,1} W_{2,1}) P_1 & -\sqrt{m} \sqrt{n_2} \lambda_{2,1} & l_n' \sqrt{n_2} (I_{n_2} - \lambda_{2,2} W_{2,2}) P_2 & 1 - \lambda_{2,2} \end{bmatrix}$$

(B.6)
|\(H'(l_N - \sum_{k=1}^{2} \sum_{i=1}^{2} \lambda_{k,i} G_{k,i})H|\) is given by

\[
\begin{vmatrix}
  P'_1(l_{n_1} - \lambda_{1,1} W_{1,1}) P_1 & -\lambda_{1,2} P'_1 W_{1,2} P_2 & 0 \\
  -\lambda_{2,1} P'_2 W_{2,1} P_1 & P'_2(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2 & 0 \\
  \frac{\ell_0}{\sqrt{\kappa_0}}(-\lambda_{2,1} W_{2,1}) P_1 & \frac{\ell_0}{\sqrt{\kappa_0}}(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2 & 1 - \lambda_{2,2}
\end{vmatrix}

+ (-1)^{n_2}(-\sqrt{\frac{n_2}{n_1}} \lambda_{2,1})
\begin{vmatrix}
  P'_1(l_{n_1} - \lambda_{1,1} W_{1,1}) P_1 & -\lambda_{1,2} P'_1 W_{1,2} P_2 & 0 \\
  -\lambda_{2,1} P'_2 W_{2,1} P_1 & P'_2(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2 & 0 \\
  \frac{\ell_0}{\sqrt{\kappa_0}}(l_{n_1} - \lambda_{1,1} W_{1,1}) P_1 & \frac{\ell_0}{\sqrt{\kappa_0}}(-\lambda_{1,2} W_{1,2}) P_2 & -\sqrt{\frac{n_1}{n_2}} \lambda_{1,2}
\end{vmatrix}

= (1 - \lambda_{1,1})(1 - \lambda_{2,2})
\begin{vmatrix}
  P'_1(l_{n_1} - \lambda_{1,11} W_{1,1}) P_1 & -\lambda_{1,2} P'_1 W_{1,2} P_2 \\
  -\lambda_{2,1} P'_2 W_{2,1} P_1 & P'_2(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2 \\
  \lambda_{1,2} P'_1(l_{n_1} - \lambda_{1,11} W_{1,1}) P_1 & -\lambda_{1,2} P'_1 W_{1,2} P_2 \\
  -\lambda_{2,1} P'_2 W_{2,1} P_1 & P'_2(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2
\end{vmatrix}

= \left( (1 - \lambda_{1,1})(1 - \lambda_{2,2}) - \lambda_{1,2}\lambda_{2,1} \right)
\begin{vmatrix}
  P'_1(l_{n_1} - \lambda_{1,11} W_{1,1}) P_1 & -\lambda_{1,2} P'_1 W_{1,2} P_2 \\
  -\lambda_{2,1} P'_2 W_{2,1} P_1 & P'_2(l_{n_2} - \lambda_{2,2} W_{2,2}) P_2
\end{vmatrix}

\text{As } \# = |J_P(l_N - \sum_{k=1}^{2} \sum_{i=1}^{2} \lambda_{k,i} G_{k,i})J_P| = |l_{N-2} - \sum_{k=1}^{2} \sum_{i=1}^{2} \lambda_{k,i} \tilde{G}_{k,i}|, \quad |l_{N-2} - \sum_{k=1}^{2} \sum_{i=1}^{2} \lambda_{k,i} \tilde{G}_{k,i}| = \\
\left( (1 - \lambda_{1,1})(1 - \lambda_{2,2}) - \lambda_{1,2}\lambda_{2,1} \right)|l_{N-2} - \sum_{k=1}^{2} \sum_{i=1}^{2} \lambda_{k,i} \tilde{G}_{k,i}|. \text{ The Lemma holds for two network group case. From this (in particular, from the matrix } \text{[B.6]}, \text{ we can easily see that the Lemma holds for any number of network case. For example, when there are three network groups, } f(\Lambda) = (1 - \lambda_{1,1})(1 - \lambda_{2,2})(1 - \lambda_{3,3}) - (1 - \lambda_{1,1})\lambda_{2,3}\lambda_{3,2} - (1 - \lambda_{2,2})\lambda_{1,3}\lambda_{3,1} - (1 - \lambda_{3,3})\lambda_{2,1}\lambda_{1,2} - \lambda_{1,3}\lambda_{2,1}\lambda_{3,2}\right)
B.4 Derivation of the best IV and the $F_i$s, and its asymptotic distribution

For the best IV and the corresponding $F_i$, we first derive the covariance matrix for the moment conditions such that $\text{var}(g(\theta)) = \Omega$ where

$$\Omega = \begin{bmatrix}
\sigma^2 Z' J_Q Z & \mu_3 Z' J_Q \omega_D \\
\mu_3 \omega_D' J_Q Z & (\mu_4 - 3\sigma^4) \omega_D' \omega_D + \sigma^4 \Delta
\end{bmatrix}$$  \hspace{1cm} (B.8)

with $\omega_D = [\text{vec}_D(J_Q F_1 J_Q)\ldots\text{vec}_D(J_Q F_m J_Q)]$, and $\Delta = [\text{vec}(J_Q F_1^\prime J_Q)\ldots\text{vec}(J_Q F_m^\prime J_Q)'' \text{vec}(J_Q F_1^\prime J_Q)\ldots \text{vec}(J_Q F_m^\prime J_Q)]$. Because, in our model, the error terms are normally distributed which implies $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$ so simply $\Omega = \begin{bmatrix}
\sigma^2 Z' J_Q Z & 0 \\
0 & \sigma^4 \Delta
\end{bmatrix}$. It should be noted that the two types of moments, linear and quadratic, are not correlated when the error are normally distributed, which allows us to choose the best linear moments and the quadratic moments independently; in other words, when we have an optimal $Z$, they are still optimal given the quadratic moments with $F_i$s.

We know that following the Hansen setting and approach (1982), the asymptotic distribution of the GMM estimator with the inverse of the variance matrix of the moment conditions, $\Omega^{-1}$, as for its distance (weighting) matrix, is given by

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{D} N(0, (\lim_{n \to \infty} \frac{1}{n} D' \Omega^{-1} D)^{-1})$$  \hspace{1cm} (B.9)

where $D = \frac{\partial E(g(\theta))}{\partial \theta}$. In our model, as $J_Q G_i Y = J_Q G_i (I_n - \sum_{i=1}^d \lambda_i G_i)^{-1} X \beta + J_Q G_i (I_n -$
\[ \sum_{i=1}^{4} \lambda_i G_i)^{-1} U, \]

\[
D = \frac{\partial E(g(\theta_0))}{\partial \theta'} \\
= \begin{bmatrix}
Z' J_Q G_1 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1} X \beta_0 & \ldots & Z' J_Q G_4 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1} X \beta_0 & Z' J_Q X \\
\sigma_0^2 tr(J_Q F_1^s J_Q G_1 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1}) & \ldots & \sigma_0^2 tr(J_Q F_1^s J_Q G_4 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1}) & 0 \\
\vdots & \ldots & \vdots & \vdots \\
\sigma_0^2 tr(J_Q F_m^s J_Q G_1 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1}) & \ldots & \sigma_0^2 tr(J_Q F_m^s J_Q G_4 (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i)^{-1}) & 0 \\
\end{bmatrix}
= \begin{bmatrix}
Z' J_Q D_{11} & Z' J_Q X \\
\sigma_0^2 D_{21} & 0 \\
\end{bmatrix}
\]

(B.10)

where \( D_{11} = [ C_1^0 X \beta_0, \ldots, C_4^0 X \beta_0 ] \) with \( C_i^0 = G_i (I_n - \sum_{i=1}^{4} \lambda_i^0 G_i) \)

and \( D_{21} = \begin{bmatrix}
\text{tr}(J_Q F_1^s J_Q C_1^0) & \ldots & \text{tr}(J_Q F_1^s J_Q C_4^0) \\
\vdots & \ldots & \vdots \\
\text{tr}(J_Q F_m^s J_Q C_1^0) & \ldots & \text{tr}(J_Q F_m^s J_Q C_4^0) \\
\end{bmatrix} \)

Then, under certain regularity conditions as in Lee (2007b) and Lee and Liu (2010), the asymptotic distribution of the estimators, \( \theta \), derived from \( \min_{\theta} g(\theta)' \hat{\Omega}^{-1} g(\theta) \), is given by

\[ \sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{1}{n} \Sigma^{-1}) \]

where \( \Sigma = D' \hat{\Omega}^{-1} D \)

\[
= \begin{bmatrix}
D_{11}' J_Q Z & \sigma_0^2 D_{21}' \\
X' J_Q Z & 0 \\
\end{bmatrix}
\begin{bmatrix}
\frac{1}{\sigma^2} (Z' J_Q Z)^{-1} & 0 \\
0 & \frac{1}{\sigma^2} \Delta^{-1} \\
\end{bmatrix}
\begin{bmatrix}
Z' J_Q D_{11} & Z' J_Q X \\
\sigma_0^2 D_{21} & 0 \\
\end{bmatrix}
\]

(B.11)

\[
\begin{array}{c}
\begin{cases}
(1) \\
(2)
\end{cases}
\end{array}
\]

We can notice that the asymptotic variance is the sum of the two parts, (1) that is affected by the choice of \( F_i \)'s, and (2) that is affected by the choice of \( Z \), and the two are independent which
is due to the normality assumption in the error terms. Then, by the generalized Schwartz inequality, the best linear IV will be \( J_Q Z_B = J_Q [D_{11}, X] = J_Q [C^0_i X \beta_0, \ldots, C^0_i X \beta_0, X] \), which reduce (2) to \( \frac{1}{\sigma^2}(Z'_B J_Q Z_B) \). And for the best quadratic moments, as \( \text{vec}'(A')\text{vec}(B^s) = \text{tr}(AB^s) = \frac{1}{2}\text{tr}(A^s B^s) = \frac{1}{2}\text{vec}'(A^s)\text{vec}(B^s) \) for any conformable matrix A and B, and
\[
\text{tr}(J_Q F^s_i J_Q C^0_i) = \text{tr}(J_Q F^s_i J_Q C^0_i) = \text{tr}(J_Q F^s_i J_Q (C^0_i - \frac{\text{tr}(C^0_i J_Q)}{n-2} J_Q) J_Q) = \frac{1}{2}\text{tr}(J_Q F^s_i J_Q (C^0_i - \frac{\text{tr}(C^0_i J_Q)}{n-2} J_Q) J_Q)
\]
\[
\frac{\text{tr}(C^0_i J_Q)}{n-2} J_Q)^s J_Q = \frac{1}{2}\text{vec}'(J_Q F^s_i J_Q)\text{vec}(J_Q (C^0_i - \frac{\text{tr}(C^0_i J_Q)}{n-2} J_Q)^s J_Q), \Delta \text{ and } D_{21} \text{ can be rewritten as }
\Delta = \frac{1}{2}[\text{vec}(J_Q F^s_i J_Q)\text{vec}(J_Q F^s_i J_Q)] [\text{vec}(J_Q F^s_i J_Q)\text{vec}(J_Q F^s_i J_Q)] \text{ and } D_{21} = \frac{1}{2}[\text{vec}(J_Q F^s_i J_Q)\text{vec}(J_Q F^s_i J_Q)] .
\]
We need to note that (3) allow us to transform the \( D_{21} \) to the form where we can use the generalized Schwartz inequality while satisfying \( \text{tr}(F_i J_Q) = 0 \) when we set \( F_i = (3) \). Then, by the generalized Schwartz inequality, the best \( F_B \) is given by
\[
F_B = (F_{B,i}) = (C^0_i - \frac{\text{tr}(C^0_i J_Q)}{n-2} J_Q) \text{ for } i = 1, \ldots, 4,
\]
which leads to the maximum value for
\[
D_{21}' \Delta^{-1} D_{21} = \frac{1}{2}[\text{vec}(J_Q F^s_{B,1} J_Q)\text{vec}(J_Q F^s_{B,4} J_Q)] [\text{vec}(J_Q F^s_{B,1} J_Q)\text{vec}(J_Q F^s_{B,4} J_Q)] = \frac{1}{2}(\text{tr}(J_Q F^s_{B,1} J_Q F^s_{B,4} J_Q)), i,j = (\text{tr}(J_Q F^s_{B,1} J_Q F^s_{B,4} J_Q)), i,j \text{ for } i,j = 1, \ldots, 4.
\]
Then, the asymptotic distribution of the Best GMM (BGMM) with the \( F_B, Z_B \) and the distance matrix of \( \Omega^{-1} \) is given by

\[
\sqrt{n}(\hat{\theta}_{BGM} - \theta_0) \sim N(0, \lim_{n \to \infty} \frac{1}{n} \Sigma)^{-1})
\]
where
\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & 0 \\
0 & 0
\end{bmatrix} + \frac{1}{\sigma^2}(Z'_B J_Q Z_B) \text{ and } \Sigma_{11} = (\text{tr}(J_Q F^s_{B,i} J_Q C^0_j))_{i,j} \text{ for } i,j = 1, \ldots, 4.
\]

(B.12)

We can show that, under the regularity conditions, this asymptotic variance is equivalent with the one from MLE (Lee, 2007b).
B.5 Descriptive Statistics for NBA Data

Descriptive statistics for each team in each season are in Table B.1 - B.3. Teams in tables are ranked based on the estimated peer-effect in Table 2.1. For each team we present the mean and standard deviations of wins produced per minute (Wins), player experience (Experience), and fatigue (Fatigue) in Tables B.1 - B.3. Experience is minutes played from the start of the game to the end of the last period, and Fatigue is minutes continuously played until the end of the last period. Each team played 82 games in each season. We also report the number of periods (Periods), observations (Observations = 5 \times Periods) and average duration per period (ADP).
### Table B.1: Season 2013-14

<table>
<thead>
<tr>
<th>Team</th>
<th>Wins (wins/min)</th>
<th>Experience (mins)</th>
<th>Fatigue (mins)</th>
<th>Games</th>
<th>Periods</th>
<th>Observations</th>
<th>ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTA</td>
<td>0.0049</td>
<td>0.0248</td>
<td>13.794</td>
<td>9.0010</td>
<td>5.2439</td>
<td>4.7098</td>
<td>82</td>
</tr>
<tr>
<td>BOS</td>
<td>0.0050</td>
<td>0.0251</td>
<td>13.860</td>
<td>9.2391</td>
<td>4.6564</td>
<td>4.3956</td>
<td>82</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0064</td>
<td>0.0257</td>
<td>13.735</td>
<td>8.9724</td>
<td>5.0906</td>
<td>4.5023</td>
<td>82</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0055</td>
<td>0.0250</td>
<td>14.594</td>
<td>9.6572</td>
<td>4.9988</td>
<td>4.6697</td>
<td>82</td>
</tr>
<tr>
<td>IND</td>
<td>0.0058</td>
<td>0.0243</td>
<td>14.287</td>
<td>9.4832</td>
<td>5.1662</td>
<td>4.8740</td>
<td>82</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0062</td>
<td>0.0257</td>
<td>13.739</td>
<td>9.4149</td>
<td>5.4394</td>
<td>5.0011</td>
<td>82</td>
</tr>
<tr>
<td>DAL</td>
<td>0.0063</td>
<td>0.0256</td>
<td>13.389</td>
<td>8.9545</td>
<td>3.6121</td>
<td>3.6704</td>
<td>82</td>
</tr>
<tr>
<td>OKC</td>
<td>0.0070</td>
<td>0.0257</td>
<td>13.146</td>
<td>9.0519</td>
<td>5.1941</td>
<td>5.0077</td>
<td>82</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0071</td>
<td>0.0258</td>
<td>12.257</td>
<td>8.1465</td>
<td>4.0379</td>
<td>3.8646</td>
<td>82</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0061</td>
<td>0.0253</td>
<td>14.235</td>
<td>9.4367</td>
<td>4.7076</td>
<td>4.3403</td>
<td>82</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0055</td>
<td>0.0237</td>
<td>12.717</td>
<td>8.5106</td>
<td>4.7137</td>
<td>4.2582</td>
<td>82</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0050</td>
<td>0.0247</td>
<td>13.731</td>
<td>9.1616</td>
<td>4.8408</td>
<td>4.5234</td>
<td>82</td>
</tr>
<tr>
<td>PHI</td>
<td>0.0049</td>
<td>0.0259</td>
<td>13.491</td>
<td>9.0089</td>
<td>4.0930</td>
<td>3.9932</td>
<td>82</td>
</tr>
<tr>
<td>NOP</td>
<td>0.0061</td>
<td>0.0253</td>
<td>13.406</td>
<td>9.0509</td>
<td>4.7968</td>
<td>4.5722</td>
<td>82</td>
</tr>
<tr>
<td>DET</td>
<td>0.0061</td>
<td>0.0254</td>
<td>14.905</td>
<td>9.6595</td>
<td>5.4940</td>
<td>5.0559</td>
<td>82</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0051</td>
<td>0.0249</td>
<td>13.876</td>
<td>9.0502</td>
<td>4.8941</td>
<td>4.6383</td>
<td>82</td>
</tr>
<tr>
<td>LAL</td>
<td>0.0058</td>
<td>0.0253</td>
<td>13.916</td>
<td>9.0277</td>
<td>5.1118</td>
<td>4.6862</td>
<td>82</td>
</tr>
<tr>
<td>NYK</td>
<td>0.0054</td>
<td>0.0243</td>
<td>14.462</td>
<td>9.4261</td>
<td>5.0768</td>
<td>4.7384</td>
<td>82</td>
</tr>
<tr>
<td>DEN</td>
<td>0.0062</td>
<td>0.0262</td>
<td>13.251</td>
<td>8.7821</td>
<td>4.8769</td>
<td>4.5826</td>
<td>82</td>
</tr>
<tr>
<td>GSW</td>
<td>0.0067</td>
<td>0.0262</td>
<td>14.840</td>
<td>9.9523</td>
<td>5.3323</td>
<td>5.5644</td>
<td>82</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0057</td>
<td>0.0239</td>
<td>13.626</td>
<td>9.0649</td>
<td>4.9568</td>
<td>4.6083</td>
<td>82</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0055</td>
<td>0.0247</td>
<td>13.286</td>
<td>8.7265</td>
<td>4.7021</td>
<td>4.4664</td>
<td>82</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0059</td>
<td>0.0251</td>
<td>14.008</td>
<td>9.3787</td>
<td>5.4816</td>
<td>5.3535</td>
<td>82</td>
</tr>
<tr>
<td>HOU</td>
<td>0.0071</td>
<td>0.0259</td>
<td>14.571</td>
<td>9.6148</td>
<td>5.3114</td>
<td>5.0911</td>
<td>82</td>
</tr>
<tr>
<td>ORL</td>
<td>0.0052</td>
<td>0.0245</td>
<td>13.862</td>
<td>9.1897</td>
<td>5.0372</td>
<td>4.7521</td>
<td>82</td>
</tr>
<tr>
<td>ATL</td>
<td>0.0058</td>
<td>0.0254</td>
<td>13.558</td>
<td>8.8973</td>
<td>4.1517</td>
<td>3.8550</td>
<td>82</td>
</tr>
<tr>
<td>LAC</td>
<td>0.0068</td>
<td>0.0263</td>
<td>14.621</td>
<td>9.8490</td>
<td>5.0150</td>
<td>4.7810</td>
<td>82</td>
</tr>
<tr>
<td>POR</td>
<td>0.0062</td>
<td>0.0255</td>
<td>14.455</td>
<td>9.5653</td>
<td>4.6440</td>
<td>4.5626</td>
<td>82</td>
</tr>
<tr>
<td>CHI</td>
<td>0.0053</td>
<td>0.0237</td>
<td>15.545</td>
<td>9.4849</td>
<td>6.5129</td>
<td>5.6085</td>
<td>82</td>
</tr>
<tr>
<td>MIA</td>
<td>0.0064</td>
<td>0.0245</td>
<td>13.522</td>
<td>8.8854</td>
<td>4.7217</td>
<td>4.3458</td>
<td>82</td>
</tr>
</tbody>
</table>

ADP: the average duration per period
<table>
<thead>
<tr>
<th>Team</th>
<th>Wins (wins/minute)</th>
<th>Experience (mins)</th>
<th>Fatigue (mins)</th>
<th>Games</th>
<th>Periods</th>
<th>Observations</th>
<th>ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>IND</td>
<td>0.0058</td>
<td>0.0256</td>
<td>13.037</td>
<td>8.2898</td>
<td>4.5181</td>
<td>4.0489</td>
<td>82</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0060</td>
<td>0.0256</td>
<td>14.016</td>
<td>9.0446</td>
<td>4.9225</td>
<td>4.3733</td>
<td>82</td>
</tr>
<tr>
<td>UTA</td>
<td>0.0059</td>
<td>0.0261</td>
<td>13.820</td>
<td>8.8752</td>
<td>3.9650</td>
<td>3.8336</td>
<td>82</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0048</td>
<td>0.0265</td>
<td>13.140</td>
<td>8.4847</td>
<td>3.6735</td>
<td>3.6020</td>
<td>82</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0054</td>
<td>0.0252</td>
<td>14.878</td>
<td>9.6813</td>
<td>5.3879</td>
<td>4.9777</td>
<td>82</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0064</td>
<td>0.0254</td>
<td>14.379</td>
<td>9.4663</td>
<td>4.8025</td>
<td>4.5130</td>
<td>82</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0059</td>
<td>0.0252</td>
<td>13.231</td>
<td>8.7861</td>
<td>4.8437</td>
<td>4.4499</td>
<td>82</td>
</tr>
<tr>
<td>LAC</td>
<td>0.0068</td>
<td>0.0258</td>
<td>14.355</td>
<td>9.6817</td>
<td>4.9482</td>
<td>4.6120</td>
<td>82</td>
</tr>
<tr>
<td>ORL</td>
<td>0.0052</td>
<td>0.0244</td>
<td>14.011</td>
<td>9.4634</td>
<td>4.7217</td>
<td>4.4842</td>
<td>82</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0057</td>
<td>0.0254</td>
<td>13.775</td>
<td>9.1679</td>
<td>4.8073</td>
<td>4.4546</td>
<td>82</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0066</td>
<td>0.0254</td>
<td>12.364</td>
<td>8.2696</td>
<td>3.7030</td>
<td>3.6057</td>
<td>82</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0058</td>
<td>0.0240</td>
<td>13.292</td>
<td>8.8446</td>
<td>4.9134</td>
<td>4.4669</td>
<td>82</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0052</td>
<td>0.0245</td>
<td>13.785</td>
<td>9.1974</td>
<td>5.0943</td>
<td>4.6681</td>
<td>82</td>
</tr>
<tr>
<td>NOP</td>
<td>0.0062</td>
<td>0.0255</td>
<td>14.415</td>
<td>9.4324</td>
<td>5.0235</td>
<td>4.6812</td>
<td>82</td>
</tr>
<tr>
<td>DEN</td>
<td>0.0056</td>
<td>0.0260</td>
<td>13.238</td>
<td>9.0719</td>
<td>4.9736</td>
<td>4.7908</td>
<td>82</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0064</td>
<td>0.0258</td>
<td>13.799</td>
<td>8.6910</td>
<td>4.5766</td>
<td>4.1246</td>
<td>82</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0055</td>
<td>0.0247</td>
<td>13.205</td>
<td>8.7309</td>
<td>4.6170</td>
<td>4.2395</td>
<td>82</td>
</tr>
<tr>
<td>DAL</td>
<td>0.0063</td>
<td>0.0265</td>
<td>13.168</td>
<td>8.7553</td>
<td>3.9218</td>
<td>3.8520</td>
<td>82</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0057</td>
<td>0.0247</td>
<td>13.076</td>
<td>8.4892</td>
<td>4.8779</td>
<td>4.3510</td>
<td>82</td>
</tr>
<tr>
<td>ATL</td>
<td>0.0062</td>
<td>0.0263</td>
<td>13.310</td>
<td>8.7269</td>
<td>3.6930</td>
<td>3.5976</td>
<td>82</td>
</tr>
<tr>
<td>HOU</td>
<td>0.0064</td>
<td>0.0263</td>
<td>14.511</td>
<td>9.3455</td>
<td>4.6074</td>
<td>4.4417</td>
<td>82</td>
</tr>
<tr>
<td>LAL</td>
<td>0.0055</td>
<td>0.0251</td>
<td>13.234</td>
<td>8.4778</td>
<td>5.2972</td>
<td>4.9816</td>
<td>82</td>
</tr>
<tr>
<td>POR</td>
<td>0.0063</td>
<td>0.0251</td>
<td>13.768</td>
<td>9.3065</td>
<td>4.3383</td>
<td>4.1583</td>
<td>82</td>
</tr>
<tr>
<td>CHI</td>
<td>0.0064</td>
<td>0.0252</td>
<td>15.114</td>
<td>9.4816</td>
<td>5.4462</td>
<td>4.9676</td>
<td>82</td>
</tr>
<tr>
<td>MIA</td>
<td>0.0054</td>
<td>0.0251</td>
<td>13.764</td>
<td>8.9606</td>
<td>4.7052</td>
<td>4.4468</td>
<td>82</td>
</tr>
<tr>
<td>DET</td>
<td>0.0056</td>
<td>0.0253</td>
<td>13.792</td>
<td>8.7494</td>
<td>5.0442</td>
<td>4.4047</td>
<td>82</td>
</tr>
<tr>
<td>OKC</td>
<td>0.0064</td>
<td>0.0265</td>
<td>13.714</td>
<td>9.2057</td>
<td>5.1044</td>
<td>4.6884</td>
<td>82</td>
</tr>
<tr>
<td>GSW</td>
<td>0.0076</td>
<td>0.0261</td>
<td>13.260</td>
<td>8.7507</td>
<td>4.7493</td>
<td>4.3759</td>
<td>82</td>
</tr>
<tr>
<td>NYK</td>
<td>0.0048</td>
<td>0.0244</td>
<td>12.849</td>
<td>8.5494</td>
<td>4.3007</td>
<td>4.1936</td>
<td>82</td>
</tr>
<tr>
<td>BOS</td>
<td>0.0055</td>
<td>0.0254</td>
<td>13.221</td>
<td>8.4947</td>
<td>4.2091</td>
<td>3.9081</td>
<td>82</td>
</tr>
</tbody>
</table>

ADP: the average duration per period
<table>
<thead>
<tr>
<th>Team</th>
<th>Wins (wins/min)</th>
<th>Experience (mins)</th>
<th>Fatigue (mins)</th>
<th>Games</th>
<th>Periods</th>
<th>Observations</th>
<th>ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>CHI</td>
<td>0.0059</td>
<td>0.0249</td>
<td>13.443</td>
<td>8.153</td>
<td>4.6841</td>
<td>4.4133</td>
<td>82</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0072</td>
<td>0.0253</td>
<td>12.260</td>
<td>8.264</td>
<td>3.8246</td>
<td>3.6986</td>
<td>82</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0058</td>
<td>0.0253</td>
<td>13.162</td>
<td>8.9522</td>
<td>4.8360</td>
<td>4.3706</td>
<td>82</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0055</td>
<td>0.0242</td>
<td>14.327</td>
<td>9.5646</td>
<td>5.5877</td>
<td>5.3570</td>
<td>82</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0056</td>
<td>0.0252</td>
<td>13.569</td>
<td>8.8102</td>
<td>4.8904</td>
<td>4.4279</td>
<td>82</td>
</tr>
<tr>
<td>ORL</td>
<td>0.0060</td>
<td>0.0249</td>
<td>13.366</td>
<td>9.1810</td>
<td>5.1182</td>
<td>4.9035</td>
<td>82</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0053</td>
<td>0.0252</td>
<td>13.583</td>
<td>8.9497</td>
<td>4.5742</td>
<td>4.2478</td>
<td>82</td>
</tr>
<tr>
<td>DET</td>
<td>0.0059</td>
<td>0.0257</td>
<td>14.666</td>
<td>9.4099</td>
<td>4.8601</td>
<td>4.5260</td>
<td>82</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0064</td>
<td>0.0261</td>
<td>14.171</td>
<td>9.0768</td>
<td>4.8797</td>
<td>4.5854</td>
<td>82</td>
</tr>
<tr>
<td>POR</td>
<td>0.0066</td>
<td>0.0261</td>
<td>13.415</td>
<td>8.7288</td>
<td>4.6371</td>
<td>4.1711</td>
<td>82</td>
</tr>
<tr>
<td>BOS</td>
<td>0.0061</td>
<td>0.0266</td>
<td>13.568</td>
<td>8.7688</td>
<td>4.5319</td>
<td>4.1359</td>
<td>82</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0068</td>
<td>0.0254</td>
<td>13.991</td>
<td>8.9852</td>
<td>4.8002</td>
<td>4.2214</td>
<td>82</td>
</tr>
<tr>
<td>UTA</td>
<td>0.0059</td>
<td>0.0256</td>
<td>13.565</td>
<td>8.9888</td>
<td>3.8109</td>
<td>3.7618</td>
<td>82</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0066</td>
<td>0.0264</td>
<td>13.731</td>
<td>9.0387</td>
<td>4.5803</td>
<td>4.2359</td>
<td>82</td>
</tr>
<tr>
<td>HOU</td>
<td>0.0063</td>
<td>0.0269</td>
<td>13.822</td>
<td>9.4650</td>
<td>4.4653</td>
<td>4.7234</td>
<td>82</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0068</td>
<td>0.0254</td>
<td>13.590</td>
<td>8.9156</td>
<td>4.9244</td>
<td>4.5149</td>
<td>82</td>
</tr>
<tr>
<td>ATL</td>
<td>0.0063</td>
<td>0.0264</td>
<td>13.145</td>
<td>8.6175</td>
<td>3.3534</td>
<td>3.3129</td>
<td>82</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0053</td>
<td>0.0261</td>
<td>13.941</td>
<td>9.4149</td>
<td>5.2119</td>
<td>5.0957</td>
<td>82</td>
</tr>
<tr>
<td>LAL</td>
<td>0.0056</td>
<td>0.0256</td>
<td>12.758</td>
<td>8.3083</td>
<td>4.9857</td>
<td>4.4327</td>
<td>82</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0061</td>
<td>0.0257</td>
<td>13.714</td>
<td>9.0765</td>
<td>4.7863</td>
<td>4.4908</td>
<td>82</td>
</tr>
<tr>
<td>NYK</td>
<td>0.0058</td>
<td>0.0251</td>
<td>13.252</td>
<td>8.9107</td>
<td>4.4921</td>
<td>4.3276</td>
<td>82</td>
</tr>
<tr>
<td>OKC</td>
<td>0.0073</td>
<td>0.0265</td>
<td>13.545</td>
<td>8.9972</td>
<td>4.5932</td>
<td>4.3073</td>
<td>82</td>
</tr>
<tr>
<td>LAC</td>
<td>0.0065</td>
<td>0.0263</td>
<td>13.064</td>
<td>8.7264</td>
<td>4.6467</td>
<td>4.2978</td>
<td>82</td>
</tr>
<tr>
<td>MIA</td>
<td>0.0065</td>
<td>0.0253</td>
<td>14.409</td>
<td>8.8739</td>
<td>4.7797</td>
<td>4.2856</td>
<td>82</td>
</tr>
<tr>
<td>NOP</td>
<td>0.0060</td>
<td>0.0252</td>
<td>14.227</td>
<td>9.1684</td>
<td>5.0658</td>
<td>4.5300</td>
<td>82</td>
</tr>
<tr>
<td>PHI</td>
<td>0.0051</td>
<td>0.0265</td>
<td>13.129</td>
<td>8.5196</td>
<td>3.6186</td>
<td>3.5643</td>
<td>82</td>
</tr>
<tr>
<td>GSW</td>
<td>0.0081</td>
<td>0.0268</td>
<td>13.157</td>
<td>9.1074</td>
<td>4.2704</td>
<td>4.0395</td>
<td>82</td>
</tr>
<tr>
<td>DAL</td>
<td>0.0058</td>
<td>0.0257</td>
<td>13.475</td>
<td>8.8809</td>
<td>4.0182</td>
<td>3.9327</td>
<td>82</td>
</tr>
<tr>
<td>IND</td>
<td>0.0061</td>
<td>0.0255</td>
<td>13.900</td>
<td>8.9143</td>
<td>4.6639</td>
<td>4.1757</td>
<td>82</td>
</tr>
<tr>
<td>DEN</td>
<td>0.0058</td>
<td>0.0257</td>
<td>13.450</td>
<td>8.7122</td>
<td>5.0957</td>
<td>4.6100</td>
<td>82</td>
</tr>
</tbody>
</table>

ADP: the average duration per period
B.6 Estimation results

The estimation results for the main equation are in Tables B.4 - B.6 for each team in each season. As the focus is the peer-effects and not the marginal effects of other important variables, we only estimate the main equation after the within transformation. Therefore, only the marginal effects for variables that vary at the player-level are identified. Teams in tables are ranked based on the estimated peer-effect in Table 2.1.
Table B.4: 2013-14 Season Estimates.

<table>
<thead>
<tr>
<th>Team</th>
<th>Parameter values</th>
<th>T statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_{rr}$</td>
<td>$\lambda_{rk}$</td>
</tr>
<tr>
<td>UTA</td>
<td>0.0446</td>
<td>0.0565</td>
</tr>
<tr>
<td>BOS</td>
<td>0.0442</td>
<td>0.0009</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0410</td>
<td>-0.0519</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0314</td>
<td>-0.0052</td>
</tr>
<tr>
<td>IND</td>
<td>0.0274</td>
<td>0.0317</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0259</td>
<td>-0.0399</td>
</tr>
<tr>
<td>DAL</td>
<td>0.0223</td>
<td>-0.0269</td>
</tr>
<tr>
<td>OKC</td>
<td>0.0208</td>
<td>-0.0119</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0206</td>
<td>0.0214</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0198</td>
<td>0.0279</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0186</td>
<td>-0.0008</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0175</td>
<td>0.0355</td>
</tr>
<tr>
<td>PHI</td>
<td>0.0171</td>
<td>0.0544</td>
</tr>
<tr>
<td>NOP</td>
<td>0.0165</td>
<td>-0.0103</td>
</tr>
<tr>
<td>DET</td>
<td>0.0150</td>
<td>-0.0376</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0142</td>
<td>-0.0351</td>
</tr>
<tr>
<td>LAL</td>
<td>0.0133</td>
<td>0.0037</td>
</tr>
<tr>
<td>NYK</td>
<td>0.0089</td>
<td>-0.0057</td>
</tr>
<tr>
<td>DEN</td>
<td>0.0089</td>
<td>-0.0072</td>
</tr>
<tr>
<td>GSW</td>
<td>0.0069</td>
<td>0.0106</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0056</td>
<td>-0.0188</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0044</td>
<td>-0.0080</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0004</td>
<td>0.0087</td>
</tr>
<tr>
<td>HOU</td>
<td>-0.0020</td>
<td>-0.0272</td>
</tr>
<tr>
<td>ORL</td>
<td>-0.0038</td>
<td>0.0320</td>
</tr>
<tr>
<td>ATL</td>
<td>-0.0052</td>
<td>0.0205</td>
</tr>
<tr>
<td>LAC</td>
<td>-0.0090</td>
<td>0.0163</td>
</tr>
<tr>
<td>POR</td>
<td>-0.0102</td>
<td>-0.0184</td>
</tr>
<tr>
<td>CHI</td>
<td>-0.0171</td>
<td>0.0057</td>
</tr>
<tr>
<td>MIA</td>
<td>-0.0193</td>
<td>-0.0214</td>
</tr>
</tbody>
</table>

$\lambda_{rr}$ is the peer-effect; $\lambda_{rk}$ is the competitor effect. Exper. is Experience.
### Table B.5: 2014-15 Season Estimates.

<table>
<thead>
<tr>
<th>Team</th>
<th>Parameter values</th>
<th>T statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_r$</td>
<td>$\lambda_k$</td>
</tr>
<tr>
<td>IND</td>
<td>0.0541</td>
<td>-0.0025</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0328</td>
<td>-0.0040</td>
</tr>
<tr>
<td>UTA</td>
<td>0.0321</td>
<td>0.0440</td>
</tr>
<tr>
<td>PHI</td>
<td>0.0232</td>
<td>-0.0017</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0204</td>
<td>0.0022</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0182</td>
<td>-0.0051</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0181</td>
<td>-0.0383</td>
</tr>
<tr>
<td>LAC</td>
<td>0.0167</td>
<td>-0.0212</td>
</tr>
<tr>
<td>ORL</td>
<td>0.0140</td>
<td>0.0209</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0107</td>
<td>0.0235</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0088</td>
<td>-0.0144</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0076</td>
<td>-0.0176</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0066</td>
<td>-0.0169</td>
</tr>
<tr>
<td>NOP</td>
<td>0.0056</td>
<td>0.0505</td>
</tr>
<tr>
<td>DEN</td>
<td>0.0021</td>
<td>-0.0202</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0017</td>
<td>-0.0103</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0014</td>
<td>0.0112</td>
</tr>
<tr>
<td>DAL</td>
<td>0.0011</td>
<td>0.0425</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0005</td>
<td>-0.0092</td>
</tr>
<tr>
<td>ATL</td>
<td>0.0000</td>
<td>0.0225</td>
</tr>
<tr>
<td>HOU</td>
<td>-0.0011</td>
<td>0.0383</td>
</tr>
<tr>
<td>LAL</td>
<td>-0.0022</td>
<td>0.0281</td>
</tr>
<tr>
<td>POR</td>
<td>-0.0031</td>
<td>0.0449</td>
</tr>
<tr>
<td>CHI</td>
<td>-0.0038</td>
<td>-0.0012</td>
</tr>
<tr>
<td>MIA</td>
<td>-0.0054</td>
<td>-0.0070</td>
</tr>
<tr>
<td>DET</td>
<td>-0.0064</td>
<td>-0.0734</td>
</tr>
<tr>
<td>OKC</td>
<td>-0.0071</td>
<td>0.0082</td>
</tr>
<tr>
<td>GSW</td>
<td>-0.0219</td>
<td>0.0569</td>
</tr>
<tr>
<td>NYK</td>
<td>-0.0254</td>
<td>-0.0028</td>
</tr>
<tr>
<td>BOS</td>
<td>-0.0261</td>
<td>-0.0067</td>
</tr>
</tbody>
</table>

$\lambda_r$ is the peer-effect; $\lambda_k$ is the competitor effect. Exper. is Experience.

<table>
<thead>
<tr>
<th>Team</th>
<th>Parameter values</th>
<th>T statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_{rr}$</td>
<td>$\lambda_{rk}$</td>
</tr>
<tr>
<td>CHI</td>
<td>0.0648</td>
<td>-0.0635</td>
</tr>
<tr>
<td>SAS</td>
<td>0.0522</td>
<td>0.0112</td>
</tr>
<tr>
<td>MIN</td>
<td>0.0489</td>
<td>0.0049</td>
</tr>
<tr>
<td>MIL</td>
<td>0.0457</td>
<td>0.0135</td>
</tr>
<tr>
<td>MEM</td>
<td>0.0452</td>
<td>0.0174</td>
</tr>
<tr>
<td>ORL</td>
<td>0.0365</td>
<td>0.0195</td>
</tr>
<tr>
<td>BKN</td>
<td>0.0350</td>
<td>0.0006</td>
</tr>
<tr>
<td>DET</td>
<td>0.0296</td>
<td>0.0107</td>
</tr>
<tr>
<td>SAC</td>
<td>0.0277</td>
<td>-0.0221</td>
</tr>
<tr>
<td>POR</td>
<td>0.0269</td>
<td>-0.0150</td>
</tr>
<tr>
<td>BOS</td>
<td>0.0218</td>
<td>-0.0008</td>
</tr>
<tr>
<td>TOR</td>
<td>0.0200</td>
<td>0.0006</td>
</tr>
<tr>
<td>UTA</td>
<td>0.0157</td>
<td>0.0008</td>
</tr>
<tr>
<td>CLE</td>
<td>0.0155</td>
<td>0.0273</td>
</tr>
<tr>
<td>HOU</td>
<td>0.0137</td>
<td>0.0060</td>
</tr>
<tr>
<td>CHA</td>
<td>0.0113</td>
<td>-0.0126</td>
</tr>
<tr>
<td>ATL</td>
<td>0.0084</td>
<td>0.0495</td>
</tr>
<tr>
<td>PHX</td>
<td>0.0037</td>
<td>-0.0365</td>
</tr>
<tr>
<td>LAL</td>
<td>0.0011</td>
<td>0.0548</td>
</tr>
<tr>
<td>WAS</td>
<td>0.0000</td>
<td>-0.0075</td>
</tr>
<tr>
<td>NYK</td>
<td>-0.0001</td>
<td>0.0506</td>
</tr>
<tr>
<td>OKC</td>
<td>-0.0020</td>
<td>-0.0506</td>
</tr>
<tr>
<td>LAC</td>
<td>-0.0045</td>
<td>0.0170</td>
</tr>
<tr>
<td>MIA</td>
<td>-0.0071</td>
<td>0.0267</td>
</tr>
<tr>
<td>NOP</td>
<td>-0.0111</td>
<td>0.0085</td>
</tr>
<tr>
<td>PHI</td>
<td>-0.0139</td>
<td>-0.0246</td>
</tr>
<tr>
<td>GSW</td>
<td>-0.0144</td>
<td>-0.0203</td>
</tr>
<tr>
<td>DAL</td>
<td>-0.0163</td>
<td>-0.0089</td>
</tr>
<tr>
<td>IND</td>
<td>-0.0175</td>
<td>-0.0005</td>
</tr>
<tr>
<td>DEN</td>
<td>-0.0282</td>
<td>-0.0123</td>
</tr>
</tbody>
</table>

$\lambda_{rr}$ is the peer-effect; $\lambda_{rk}$ is the competator effect. Exper. is Experience.
Appendix C

Adaptive LASSO for Stochastic Frontier Models with Many Efficient Firms

C.1 Proofs of the lemmas and the theorem

We denote

□ LSDV estimates for individual intercepts: \( \hat{\alpha}_i(0) \) for \( i = 1, \ldots, N \)

□ LASSO estimates for individual intercepts: \( \hat{\alpha}_i(\Pi) = \hat{\alpha}(\Pi) - \hat{u}(\Pi)_i \) for \( i = 1, \ldots, N \)

Lemma 3.4.1 Under the assumption 3.4.1 we have \( E((\hat{\alpha}(0) - \alpha_0)^2) = O(\frac{(\log N)^2}{T}) \), \( E(||\hat{\beta}(0) - \beta_0||^2) = O(\frac{1}{NT}) \), and \( E((\hat{u}_i(0) - u_{0,i})^2) = O(\frac{(\log N)^2}{T}) \) for \( \forall i \) as \( (N, T) \to \infty \)

Proof \( E(||\hat{\beta}(0) - \beta_0||^2) = O(\frac{1}{NT}) \) can be easily proved as it is a standard fixed effect panel data model problem in the literature. We only prove \( E((\hat{\alpha}(0) - \alpha_0)^2) = O(\frac{(\log N)^2}{T}) \) and \( E((\hat{u}_i(0) - u_{0,i})^2) = O(\frac{(\log N)^2}{T}) \) here.

It is easy to show \( (\hat{\alpha}_i(0) - \alpha_{0,i})^2 = O(\frac{1}{T}) \) for \( \forall i \) from \( \hat{\alpha}(0), = \frac{1}{T} \sum_{t=1}^{T} (y_{it} - x_{it}^\prime \hat{\beta}(0)) \) which
implies
\[ \hat{\alpha}_i(0) - \alpha_{0,i} = \frac{1}{T} \sum_{t=1}^{T} x'_t (\beta_0 - \hat{\beta}(0)) + \frac{1}{T} \sum_{t=1}^{T} v_t \Rightarrow E((\hat{\alpha}_i(0) - \alpha_{0,i})^2) = O\left(\frac{1}{T}\right) \quad (C.1) \]
due to \( E(||\hat{\beta}(0) - \beta_0||_2^2) = O\left(\frac{1}{NT}\right) \). Then, using the results from Theorem 3.2 (ii) of Park, Schmidt and Simar (1998), we can show
\[ \hat{\alpha}(0) - \alpha_0 = O_p\left(\frac{\log N}{\sqrt{T}}\right) \Rightarrow E((\hat{\alpha}(0) - \alpha_0)^2) = O\left(\frac{(\log N)^2}{T}\right) \quad (C.2) \]
which is due to \( E_{\max} \left| \frac{1}{\sqrt{T}} \sum_i v_t \right| = O(\log N) \). Next, we derive the convergence rate of \( \hat{u}_i(0) \).
By definition, \( \hat{u}_i(0) = \hat{\alpha}(0) - \hat{\alpha}_i(0) \). Then, from above results, we have
\[ \begin{align*}
\hat{u}_i(0) &= \alpha_0 + O_p\left(\frac{\log N}{\sqrt{T}}\right) - \left(\alpha_{0,i} + O_p\left(\frac{1}{\sqrt{T}}\right)\right) \Rightarrow \sup |\hat{u}_i(0) - (\alpha_0 - \alpha_{0,i})| = O_p\left(\frac{\log N}{\sqrt{T}}\right) \\
&\Rightarrow E((\hat{u}_i(0) - u_{0,i})^2) = O\left(\frac{(\log N)^2}{T}\right) \quad (C.3)
\end{align*} \]
The lemma is proved. ■

**Lemma 3.4.2** Under the assumption 3.4.1 and 3.4.2, we have (i) \( E((\hat{\alpha}(\Pi) - \alpha_0)^2) = O(\frac{1}{\delta_0 NT}) \), (ii) \( E(||\hat{\beta}_A(\Lambda) - \beta_{0,A}||_2^2) = O(\frac{1}{NT}) \), (iii) \( E((\hat{u}_{A,i}(\Pi) - u_{0,A,i})^2) = O\left(\frac{1}{T}\right) \) for \( \forall \) \( i \) \( (N, T) \to \infty \)

**Proof** (ii) has been proved in the literature. It can be proved similarly using the first step equation of (3.10). To prove (i), we use the coordinate decent algorithm in 3.3.1. The algorithm implies that \( \hat{\alpha}(\Pi) \) is estimated as a common intercept of the firms categorized as fully efficient by LASSO technique. Therefore, if we know the true model, \( \hat{\alpha}(\Pi) \) is given by \( \hat{\alpha}(\Pi) = \frac{1}{\delta_0 NT} l'_{NT,\delta} (Y - X_{1,\Lambda} \hat{\beta}_A(\Lambda)) \) where \( l_{NT,\delta} \) a vector with ones at the observations with zero inefficiency and zero at the others. From the model assumption and \( E(||\hat{\beta}_A(\Lambda) - \beta_{0,A}||_2^2) = \)
\( O_p\left(\frac{1}{\sqrt{T}}\right) \), we have

\[
(\hat{\alpha}(\Pi) - \alpha_0) = \frac{1}{\delta_0 N T} l_{NT, \delta} X_{1, A} (\beta_{0, A} - \hat{\beta}_A (\Lambda)) + \frac{1}{\delta_0 N T} l'_{NT, \delta} \nu \Rightarrow E (\hat{\alpha}(\Pi) - \alpha_0)^2 = O \left( \frac{1}{\delta_0 N T} \right)
\]

(C.4)

Next, we prove (iii). The algorithm gives us

\[
\hat{u}_{A, i}(\Pi) = \hat{\alpha}(\Pi) - \hat{\alpha}_i (0) - \frac{\Pi \hat{\pi}_i}{2 T} = \alpha_0 + O_p \left( \frac{1}{\sqrt{\delta_0 N T}} \right) - \left( \alpha_{0, i} + O_p \left( \frac{1}{\sqrt{T}} \right) \right) - \frac{\Pi \hat{\pi}_i}{2 T}
\]

(C.5)

which implies

\[
\sup \left| \hat{u}_{A, i}(\Pi) - u_{0, i} \right| = O_p \left( \frac{1}{\sqrt{T}} \right) - \frac{\Pi \hat{\pi}_i}{2 T} \leq O_p \left( \frac{1}{\sqrt{T}} \right) - O_p \left( \frac{1}{2 \sqrt{T \log N}} \right) \frac{\Pi}{\sqrt{T}} \eta^{-\gamma_u} \left( \frac{\hat{\eta}}{\eta} \right)^{-\gamma_u}
\]

(C.6)

where \( \eta = \min_i \left| u_{0, i} \right| \) and \( \hat{\eta} = \min_i \left| \hat{u}_i (\Pi) \right| \) and \( T^* = \frac{T}{(\log N) \tau} \). We show that

\[
E \left[ \left( \frac{\hat{\eta}}{\eta} \right)^2 \right] = E \left[ \left( \frac{\hat{\eta} - \eta + \eta}{\eta} \right)^2 \right] \leq \frac{2}{\eta^2} E \left[ (\hat{\eta} - \eta)^2 \right] + 2
\]

\[
\leq \frac{2}{\eta^2} E \left[ |\hat{u}_a (0) - u_{0, a}|^2 \right] + 2 E \left[ |\hat{u}_b (0) - u_{0, b}|^2 \right] + 2 = \frac{2}{\eta^2} O \left( \frac{(\log N)^2}{T} \right) + 2 = O(1)
\]

(C.7)

where \( a \) and \( b \) is defined from \( \eta = |u_{0, a}| \) and \( \hat{\eta} = |\hat{u}_b (\Pi)| \). The second inequality is straightforward if \( a = b \). Even if not, when \( \hat{\eta} > \eta \), \( E \left[ (\hat{\eta} - \eta)^2 \right] < E \left[ |\hat{u}_a (0) - u_{0, a}|^2 \right] \) and when \( \hat{\eta} < \eta \), \( E \left[ (\hat{\eta} - \eta)^2 \right] < E \left[ |\hat{u}_b (0) - u_{0, b}|^2 \right] \) so the inequality holds. The next equality is due to the fact that the convergence rate for the individual LSDV estimator is \( O_p \left( \frac{(\log N)^2}{T} \right) \). The last equality holds because

\[
O \left( \frac{(\log N)^2}{T} \frac{1}{\eta^2} \right) = O \left( \frac{\Pi}{\sqrt{T^*}} \eta^{-\gamma_u} (\log N \cdot \frac{\Pi}{\sqrt{T^*}} \cdot T^{* \gamma_u/2})^{-2/\gamma_u} \right) = o(1)
\]

(C.8)
due to the assumption 3.4.2. Therefore, (C.7) gives us

\[ \sup |\hat{u}_{A,i}(\Pi) - u_{0,i}| \leq O_p\left( \frac{1}{\sqrt{T}} \right) - O_p\left( \frac{1}{\sqrt{T \log N}} \right) o_p(1) = O_p\left( \frac{1}{\sqrt{T}} \right) \]  

(C.9)

which implies \( E((\hat{u}_{A,i}(\Pi) - u_{0,A,i})^2) = O(\frac{1}{T}) \) for \( \forall i \)

**Lemma 3.4.2** B Under the assumption 3.4.1 3.4.2 and 3.4.3 the adaptive LASSO estimator for \( \beta_0, \hat{\beta}(\Lambda) \), has the oracle property; that is, the estimator satisfy:

1. Consistency in selection: \( Pr(\{j : \hat{\beta}(\Lambda)_j \neq 0\} = A) \rightarrow 1 \)

2. Asymptotic normality: \( \sqrt{NT}(\hat{\beta}(\Lambda) - \beta_0) \rightarrow_d N(0, \sigma^2 \Sigma_{-1,A}) \)

**Proof** See Zou (2006) for proof. Only difference is the regressors in our problem involves a within transformation.

**Lemma 3.4.3** Under the assumption 3.4.1 and 3.4.2 \([\hat{\alpha}(\Pi), (\hat{\beta}_A(\Lambda)', 0'), (\hat{U}_A(\Pi)', 0')]' \) is the solution to the minimization problem of (3.3) w.p.a 1 \((N, T) \rightarrow \infty \)

**Proof** By the lemma 3.4.2 B, the selection consistency of \( \hat{\beta}(\Lambda) \) is already proved. To show \((\hat{U}_A(\Pi)', 0') \) is the solution to the minimization problem of (3.3) w.p.a 1, we need to show it satisfies \( \Psi_j \equiv P\left\{ \text{For } j \in A^c, |2x_j'(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| \leq \Pi \cdot \hat{\pi}_j \right\} \rightarrow 1 \)

or equivalently,

\[ \Psi_j \equiv P\left\{ \text{For } j \in A^c, |2x_j'(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \Pi \cdot \hat{\pi}_j \right\} \rightarrow 0 \]  

(C.10)

\(^1\text{Zou and Zhang (2009) considers the selection consistency for all the coefficients at the same time, but in our problem, even though there is an infinite number of inefficiency terms in the limit, they are independent due to the independency between the dummy variables, therefore it is suffice to consider of the selection consistency individually.} \)
With some set theorems, we have

\[
\psi_j \equiv P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j \right\}
\]

\[
= P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j \text{ and } \hat{\eta} > \eta/2 \right\}
\]

\[
+ P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j \text{ and } \hat{\eta} \leq \eta/2 \right\}
\]

\[
\leq P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j, \hat{\eta} > \eta/2 \right\} + P \left\{ \hat{\eta} \leq \eta/2 \right\}
\]

As in [C.7], a and b is defined from \( \eta = |u_{0,a}| \) and \( \hat{\eta} = |\hat{u}_b(\Pi)| \), then, we can show

\[
P \left\{ \hat{\eta} \leq \eta/2 \right\} \leq P \left\{ |\hat{u}_a(0) - u_{0,a}| \geq \eta/2 \right\} + P \left\{ |\hat{u}_b(0) - u_{0,b}| \geq \eta/2 \right\}
\]

\[
\leq E(|\hat{u}_a(0) - u_{0,a}|^2) + E(|\hat{u}_b(0) - u_{0,b}|^2)
\]

\[
\leq \frac{(\log N)^2}{\eta^2/4} = O \left( \frac{\log N}{\eta} \right), \quad \frac{1}{\eta^2}
\]

w.p.a 1. The first inequality is due to that if \( a = b = e^* \), \( \hat{\eta} \leq \eta/2 \) implies \( |\hat{\eta} - \eta| = |\hat{u}_{e^*}(0) - u_{0,e^*}| \geq \eta/2 \). Similarly, if \( a \neq b \), \( \hat{\eta} \leq \eta/2 \) implies \( |\hat{\eta} - u_{0,b}| \geq |\hat{\eta} - \eta| \geq \eta/2 \) due to \( |u_{0,b}| > \eta \) by definition of \( \eta \). So the inequality holds. Let \( M = \left( \frac{n}{T/\log N} \right)^{1/(1+\gamma_u)} \), then, we can show

\[
P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j, \hat{\eta} > \eta/2 \right\}
\]

\[
\leq P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot \hat{\pi}_j, \hat{\eta} > \eta/2, |\hat{u}_j(0)| \leq M \right\} + P \left\{ |\hat{u}_j(0)| > M \right\}
\]

\[
\leq P \left\{ |2x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))| > \pi \cdot M^{-\gamma_u}, \hat{\eta} > \eta/2 \right\} + P \left\{ |\hat{u}_j(0)| > M \right\}
\]

\[
\leq 4 \frac{M^{2\gamma_u}}{\pi^2} \cdot E \left[ |x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2 \cdot I_{\hat{\eta} > \eta/2} \right] + \frac{1}{M^2} \cdot E \left[ |\hat{u}_j(0)|^2 \right]
\]

\[
\leq 4 \frac{M^{2\gamma_u}}{\pi^2} \cdot E \left[ |x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2 \cdot I_{\hat{\eta} > \eta/2} \right] + \frac{1}{M^2} \cdot E \left( |\hat{u}_j(0) - u_{0,j}|^2 \right) \quad \text{due to} \quad u_{0,j} \in A^c
\]

\[
= 4 \frac{M^{2\gamma_u}}{\pi^2} \cdot E \left[ |x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2 \cdot I_{\hat{\eta} > \eta/2} \right] + \frac{1}{M^2} \cdot O \left( \frac{\log N}{T} \right) \quad \text{w.p.a 1 by the lemma 3.4.1}
\]

(C.13)
Next, we derive a upper bound of \( A \). By the model assumption, we have

\[
E\left[|x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2\right] = E\left[|x'_j(X_A \cdot \theta_0 + \nu - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2\right] \tag{C.14}
\]

\[
\leq 2 \cdot E\left[|x'_j X_A(\theta_0 - \hat{\theta}_A(\Lambda, \Pi))|^2\right] + 2 \cdot E\left[|x'_j \nu|^2\right]
\]

Due to \( j \in A^c \), there’s no inefficiency estimate and parameter in this firm, therefore, we have

\[
\leq 2 \cdot E\left(|x'_j X_A(\beta_0 - \hat{\beta}_A(\Lambda)) + x'_j l_{NT}(a_0 - \hat{\alpha}(\Pi))|^2\right) + 2 \cdot T \cdot \sigma^2
\]

\[
\leq 4T \cdot E\left(||X_A(\beta_0 - \hat{\beta}_A(\Lambda)||_2^2\right) + 4 \cdot T^2 \cdot E\left(||a_0 - \hat{\alpha}(\Pi)||_2^2\right) + 2 \cdot T \cdot \sigma^2 \tag{C.15}
\]

\[
\leq 4 \cdot BNT^2 \cdot E\left(||\beta_0 - \hat{\beta}_A(\Lambda)||_2^2\right) + 4 \cdot T^2 \cdot E\left(||a_0 - \hat{\alpha}(\Pi)||_2^2\right) + 2 \cdot T \cdot \sigma^2
\]

\[
\leq 4B \cdot NT^2 \cdot \{E\left(||a_0 - \hat{\alpha}(\Pi)||_2^2\right) + E\left(||\beta_0 - \hat{\beta}_A(\Lambda)||_2^2\right)\} + 2 \cdot T \cdot \sigma^2
\]

Then, \( \text{C.14} \) and \( \text{C.15} \) give us the inequality

\[
E\left[\sum_{j \in A^c} |x'_j(Y - X_A \cdot \hat{\theta}_A(\Lambda, \Pi))|^2 \cdot I_{\hat{\eta} > \eta/2}\right] \tag{C.16}
\]

\[
\leq 4BNT^2 \cdot \\{E\left(||a_0 - \hat{\alpha}(\Pi)||_2^2 \cdot I_{\hat{\eta} > \eta/2}\right) + E\left(||\beta_0 - \hat{\beta}_A(\Lambda)||_2^2 \cdot I_{\hat{\eta} > \eta/2}\right)\} + 2T \cdot \sigma^2
\]

Combining the above results and by the lemma 3.4.2, we have the upper bound of \( \Psi_j \) as

\[
\Psi_j \leq \frac{4}{M^2} \cdot \frac{\gamma_u}{\Pi^2} \cdot \left(4BNT^2 \cdot O\left(\frac{1}{NT}\right) + 2T \sigma^2\right) + \frac{1}{M^2} \cdot O\left(\frac{(\log N)^2}{T}\right) + O\left(\frac{(\log N)^2}{T}\right) \cdot \frac{1}{\eta^2}
\]

\[
\equiv \Psi_{j,1} + \Psi_{j,2} + \Psi_{j,3} \tag{C.17}
\]
w.p.a 1. Under the assumption 3.4.1 and 3.4.2, and due to C.8 we can show

\[ \Psi_{j,1} = O \left( \frac{M^2\gamma_u}{\eta^2} \cdot T \right) = O \left( \left( \frac{\eta}{\sqrt{T}} \cdot T^{*\gamma_u/2} \right)^{-2/(1+\gamma_u)} \right) \rightarrow 0 \]

\[ \Psi_{j,2} = O \left( \frac{1}{M^2} \cdot \frac{1}{T^*} \right) = O \left( \left( \frac{\eta}{\sqrt{T}} \cdot T^{*\gamma_u/2} \right)^{-2/(1+\gamma_u)} \right) \rightarrow 0 \]  

(C.18)

Thus, the proof is complete ■

**Theorem (Oracle Property)** Under the assumption 3.4.1 3.4.2 and 3.4.3, the LASSO estimator for \( \theta_0, \hat{\theta}(\Lambda, \Pi) \), has the oracle property; that is, the estimator satisfy:

1. Consistency in selection: \( Pr(\{ j : \hat{\theta}(\Lambda, \Pi)_j \neq 0 \} = A) \rightarrow 1 \) as \( (N, T) \rightarrow \infty \)

2. Asymptotic normality

1) \( \sqrt{NT}(\hat{\beta}_A(\Lambda) - \beta_{0,A}) \rightarrow_d N(0, \sigma^2_{\nu,0} \Sigma_{A,1}) \)

2) \( \sqrt{\delta_0 NT}(\hat{\alpha}(\Pi) - \alpha_0) \rightarrow_d N(0, \sigma^2_{\nu,0} \Sigma_\alpha) \)

3) \( \sqrt{T}(\hat{u}_{i,A}(\Pi) - u_{i,0}) \rightarrow_d N(0, \sigma^2_{\nu,0}) \)

w.p.a 1 as \( (N, T) \rightarrow \infty \) where \( \Sigma_\alpha = 1 + \delta_0 \frac{\gamma_{\nu_{1,A}}(X_{\nu_{1,A}}^X X_{\nu_{1,A}}^L)}{\delta_0 NT} \cdot \frac{X_{\nu_{1,A}}^L X_{\nu_{1,A}}^L}{\delta_0 NT} \).

**Proof** Using Lemma 3.4.2 B and 3.4.3 we only need to show \( P\{ \min_{i \in A} \hat{u}_i(\Pi) > 0 \} \rightarrow 1 \). We can show that

\[ \hat{u}_a(\Pi) > \min_{i \in A} u_{0,i} - |\hat{u}_a(\Pi) - u_{0,a}| \]  

(C.19)

where \( a \) is the index for \( \hat{u}_a(\Pi) = \min_{i \in A} \hat{u}_i(\Pi) \). The inequality is straightforward if \( \hat{u}(\Pi)_a > \min_{i \in A} u_{0,i} \). When \( \hat{u}_a(\Pi) < \min_{i \in A} u_{0,i} \), it implies \( \hat{u}_a(\Pi) < \min_{i \in A} u_{0,i} < u_{0,a} \), which in turn
implies \( \hat{u}_a(\Pi) - \min_{i \in A} u_{0,i} > -|\hat{u}_a(\Pi) - u_{0,a}| \). Then, due to the Lemma 3.4.2, we have

\[
\min_{j \in A} \hat{u}(\Pi)_j > \eta - O_p\left(\sqrt{\frac{1}{T}}\right) \tag{C.20}
\]

The proof is complete because we have already shown that \( \sqrt{\frac{(\log N)^2}{T}} \) converges to zero faster than \( \eta \) by C.8.

Next, we prove the asymptotic normality. From the selection consistency, we only have nonzero coefficients now. Then, each estimator's asymptotic distribution will be given as follows:

First, we already showed that \( \sqrt{N T} (\hat{\beta}_A(\Lambda) - \beta_{0,A}) \rightarrow_d N(0, \sigma^2_{v_0} \Sigma^{-1}_{A,1}) \) from Lemma 3.4.2 B.

Second, as we've seen in the Lemma 3.4.2, we have

\[
(\hat{\alpha}(\Pi) - \alpha_0) = \frac{1}{\delta_0 N T} \left( \mathbf{l}'_\delta (Y - X_{1,A} \hat{\beta}_A(\Lambda)) \right).
\]

As \((N, T) \rightarrow \infty\), we will have \( \hat{\beta}_A(\Lambda) = (X'_{1,A} Q X_{1,A})^{-1} X'_{1,A} Q Y \) w.p.a 1. Then, we can show

\[
\sqrt{\delta_0 N T} (\hat{\alpha}(\Pi) - \alpha_0) = \frac{1}{\sqrt{\delta_0 N T}} \mathbf{l}'_\delta (v - X_{1,A} (X'_{1,A} Q X_{1,A})^{-1} X'_{1,A} Q v) \tag{C.21}
\]

where we omit the inefficiency terms as they will be removed eventually. By CLT, we can show

\[
\sqrt{\delta_0 N T} (\hat{\alpha}(\Pi) - \alpha_0) \rightarrow_d N(0, \sigma^2_{v_0} \Sigma_\alpha) \tag{C.22}
\]

where

\[
\left( \frac{1}{\sqrt{\delta_0 N T}} \mathbf{l}'_\delta - \sqrt{\delta_0 N T} \left( \frac{X'_{1,A} Q X_{1,A}}{N T} \right)^{-1} \frac{1}{\sqrt{N T}} X'_{1,A} Q \right) \left( \frac{1}{\sqrt{\delta_0 N T}} \mathbf{l}'_\delta - \sqrt{\delta_0 N T} \left( \frac{X'_{1,A} Q X_{1,A}}{N T} \right)^{-1} \frac{1}{\sqrt{N T}} X'_{1,A} Q \right)' \rightarrow_p \Sigma_\alpha.
\]

This reduces to \( 1 + \delta_0 \frac{1}{\delta_0 N T} (X'_{1,A} Q X_{1,A})^{-1} X'_{1,A} \delta_0 \rightarrow_p \Sigma_\alpha \) due to \( Q l_\delta = 0 \).

Lastly, for each nonzero inefficiencies, we have F.O.C as

\[
\sum^i (y_{it,A} - \hat{\alpha}(\Pi) - x_{it,1,A} \cdot \hat{\beta}_A(\Lambda) + \hat{u}_{i,A}(\Pi)) + \frac{\Pi}{2} \cdot \hat{\pi}_i = 0 \quad \text{for } \hat{u}_{i,A}(\Pi) > 0 \tag{C.23}
\]
w.p.a 1. This leads to

\[
\sqrt{T}(\hat{u}_{i,A}(\Pi) - u_{i,0}) = \sqrt{T}(\hat{\alpha}(\Pi) - \alpha_0) + \frac{1}{\sqrt{T}} \sum_{i}^{\text{(1)}} x_{it,1,A}(\beta_A(\Lambda) - \beta_{A,0}) - \frac{1}{\sqrt{T}} \sum_{i}^{\text{(2)}} v_{it} - \frac{\Pi}{2\sqrt{T}} \cdot \hat{\pi}_i
\]

As (1) = \text{o}_p(1) and (2) = \text{o}_p(1), by CLT, we have

\[
\sqrt{T}(\hat{u}_{i,A}(\Pi) - u_{i,0}) \to_d N(0, \sigma^2_{i,0})
\]  

(C.24)

w.p.a 1 as \((N, T) \to \infty\). This completes the proof. 

C.2 The equivalence between the one step estimation of \((3.3)\) and the two step estimation of \((3.5)\) and \((3.6)\)

In order to show the equivalence between the one step estimation of \((3.3)\) and the two step estimation of \((3.5)\) and \((3.6)\), we first derive the F.O.C for \(\hat{U}(\Pi)\) in \((3.3)\), which gives us the close form solution for \(\hat{U}(\Pi)\) such that

\[
\hat{U}(\Pi) = \left[ (X_2'X_2)^{-1} \left( -X_2'(Y - \alpha \cdot l_{NT} - X_1\beta) - \Pi \cdot i_N(\hat{\pi}_i^2) \right) \right]_+
\]  

(C.26)

where \((x)_+ = x\) if \(x \geq 0\) and \(= 0\) otherwise and \(i_N(\hat{\pi}_i)\) is \(N \times 1\) vector with \(\hat{\pi}_i\) for \(i^{th}\) element.

Plugging this result into \((3.3)\) give us the equation

\[
\left\{ Q \cdot (Y - X_1\beta) - X_2 \Pi \cdot i_N(\frac{\hat{\pi}_i}{2T}) \right\} + \sum_{j=1}^{p} \lambda_j |\beta_j| + \sum_{k=1}^{N} \hat{\pi}_k |u_k|
\]  

(C.27)
where $\epsilon_i$ is for removing the positive constraints on the inefficiency estimates so $\beta$ has no corresponding constraint in (C.27). The equality is due to $Q \cdot l_{NT} = 0$. Next we derive the F.O.C for $\beta$ in (C.27), which is given by

$$-2 \cdot X_1'Q \left\{ Q \cdot (Y - X_1\beta) - X_2\Pi \cdot i_N \left( \frac{\hat{\pi}_i}{T} + \epsilon_i \right) \right\} + \Lambda \cdot i_p(\hat{\lambda}_i)$$

$$= -2 \cdot X_1'Q (Y - X_1\beta) + \Lambda \cdot i_p(\hat{\lambda}_i)$$

This equality is due to $Q \cdot Q = Q$ and $Q \cdot X_2 = 0$. This is the same with the F.O.C for $\beta$ in (3.5) so we show the equivalence between the two. ■
C.3 The proximity of the new algorithm and the standard constrained optimization algorithm

The basic setup here is the same with that in the simulation exercise. We compare the series of estimation results on $\hat{U}(\Pi)$ (RMSE, $Pr_{\hat{U}_a}, \hat{\alpha}$, Sum of squared error) from the new algorithm (NA) and one of the standard constrained algorithm (SA) (SQP in Matlab) in the 20 replications of two models: (1) $(N,T)=(20,10), \sigma_u = 1$, and (2) $(N,T)=(10,20) \sigma_u = 2$.

Figure C.1 is from model (1) and Figure C.2 is from model (2). The black dashed line is from NA and the red dotted line is from SA. In model (1), there are one or two replications showing the two algorithms produced slightly different results, but the differences are small enough to conclude that the two algorithms generally produce very similar results. In model (2), the two produced almost identical results in every replication. The new algorithm is incomparably faster than SA in estimation. For one replication, the new algorithm took on average 1.5 second whereas the standard algorithm took 345 seconds. This difference will be pronounced as the sample size increases.

**Figure C.1: Model (1)**
C.4 New algorithm

□ Using \( \beta(\Lambda) \) from the first step, compute \( \hat{\alpha}_i = \frac{1}{T} \sum_t y_{it} - x'_{it} \beta(\Lambda) \) and \( \hat{u}_i = \max_{j=1}^N \hat{\alpha}_j - \hat{\alpha}_i \) for all \( i \). Let \( \hat{\alpha}_{[1]} \leq \hat{\alpha}_{[2]} \leq \ldots \leq \hat{\alpha}_{[N]} \) be the rankings of the \( \hat{\alpha}_i \), so \( \hat{\alpha}_{[N]} = \max_{j=1}^N \hat{\alpha}_j \). Similarly, let \( \hat{u}_{[N]} \leq \hat{u}_{[N-1]} \leq \ldots \leq \hat{u}_{[1]} \) be the rankings of the \( \hat{u}_i \), so \( \hat{u}_{[N]} = \min_{j=1}^N \hat{u}_j \). We set the initial value for \( \alpha \) as \( \hat{\alpha}_{[N]} \). Denote the current values for \( \hat{u}_{[i]} \) and \( \hat{\alpha}_{[i]} \) as \( \hat{u}^{(0)}_{[i]} \) and \( \hat{\alpha}^{(0)}_{[i]} \), then we have below initial settings. Note that as \( \hat{\alpha}^{(0)} = \hat{\alpha}_{[N]} \), we have one fully efficient firm, \( \hat{u}^{(0)}_{[N]} = 0 \), now.

\[
\hat{\alpha}^{(0)}_{[N]} = \hat{\alpha}_{[N]}, \quad \hat{u}^{(0)}_{[N]} = 0
\]

□ For a given \( \Pi \), check the KKT condition for the second best firm (\( i = [N-1] \)), that is, check the sign of \( \Delta_{[N-1]} = \hat{u}^{(0)}_{[N-1]} - \Pi \frac{2[N-1]}{2T} \)

1. \( \Delta_{[N-1]} \leq 0 \), update \( \hat{u}^{(0)}_{[N-1]} \) as \( \hat{u}_{[N-1]} = 0 \), and update \( \hat{\alpha}^{(0)} \) as \( \hat{\alpha} = \frac{1}{2}(\hat{\alpha}_{[N]} + \hat{\alpha}_{[N-1]}) \).

As we have a new frontier (=\( \hat{\alpha} \)), we update the rest of the inefficiencies (from \( [N-2] \) to \([1]\)) as below.
(a) Check the KKT condition for the third best firm ($\hat{u}_{N-2}$), and if $\Delta_{N-2} \leq 0$, update $\hat{u}_{N-2}^{(0)}$ as $\hat{u}_{N-2} = 0$, and update $\hat{\alpha}^{(0)}$ as $\hat{\alpha} = \frac{1}{2}(\hat{\alpha}_{N} + \hat{\alpha}_{N-1} + \hat{\alpha}_{N-2})$, and update the rest of the inefficiencies (from $[N - 3]$ to $[1]$) as below.

(b) if $\Delta_{N-2} > 0$, do some minor updating for $\hat{\alpha}$ and $\hat{u}_{i}$ for $i = N - 2, N - 3, ..., 1$ as 2. (b) of the algorithm in 3.3.1 and report the results.
### C.5 Simulation results for $\beta$

#### Table C.1: Simulation results for $\beta$

<table>
<thead>
<tr>
<th>$(N, T, \sigma_u)$</th>
<th>Estimation accuracy: RMSE</th>
<th>Selection accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LASSO</td>
<td>Oracle</td>
</tr>
<tr>
<td>(100,10,1)</td>
<td>0.023 (0.011)</td>
<td>0.013 (0.002)</td>
</tr>
<tr>
<td>(100,50,1)</td>
<td>0.009 (0.004)</td>
<td>0.005 (0.001)</td>
</tr>
<tr>
<td>(200,50,4)</td>
<td>0.006 (0.003)</td>
<td>0.004 (0.000)</td>
</tr>
</tbody>
</table>

The numbers in main entries are the average RMSE over 1,000 replications and $|\hat{\beta}_A|$ represents the average number of nonzero estimate for $\beta_A$ (the true = 3) over the replications, and $|\hat{\beta}_{Ac}|$ represents the average number of zero estimate for $\beta_{Ac}$ (the true = 5). The corresponding variances of each measures are in next row in parentheses. The oracle is the RMSE calculated from when we know the true model and apply LSDV estimation to the model.
Vita

Hyunseok Jung

Updated: 4/10/18

Syracuse University
Department of Economics
110 Eggers Hall
Syracuse University, Syracuse, NY, 13244
Phone: +1 (315) 800-2258
Office: +1 (315) 443-9054
Email: hjung02@syr.edu
Homepage: https://hyunseokjung.weebly.com

Education

□ Ph.D., Economics, Syracuse University, 2018 (expected).
□ M.A., Public Administration, Syracuse University, 2014.
□ M.A., Development Policy, Korea Development Institute, 2013
□ B.A., Agricultural Economics, Seoul National University, 2008.

Fields of Interest

□ Primary: Theoretical & Applied Econometrics (Production & Efficiency Analysis; Spatial & Network Econometrics; Big-data/Large-dimensional Models & Shrinkage Estimation)

□ Secondary: International Trade, Microeconomic Theory

Working Papers

1. Adaptive LASSO for Stochastic Frontier Models with Many Efficient Firms

2. Network Competition and Team Chemistry in the NBA (with William Horrace and Shane Sanders)

3. Police Officer Experience and Racial Bias in Traffic Stops (with William Horrace and Shawn Rohlin, Submitted)

4. Stochastic Frontier Models with Network Selectivity (with William Horrace, R&R)
In Progress

1. Model Selection via SSE Ratios (with Yoonseok Lee)

2. Social Interactions in NYC Elementary Schools: Multiple Reference Groups Approach (with William Horrace and Jonathan Presler)

3. Peer Effects in Policing (with William Horrace and Shawn Rohlin)

4. Crowding In or Out: the Structure of Market Competition and Cooperation in a Grocery Store Chain (with Dinesh Gauri)

Presentations

□ 2018: FSU, Nevada(Reno), Arkansas(Fayetteville), 10th NAPW (scheduled)

□ 2017: 15th EWEPA(London), Albany, RPI, Binghamton, Union

Experience

□ Research: Research Associate for William C. Horrace (Aug. 2016 - Present)


□ Government: Deputy Director, Ministry of Land and Transport, Korea (2008 - 2012)

□ Directing: Cinema Academy, Korea (Directing a Short Film “Blue Frog", 2007)


Awards & Fellowships

□ W.E.B. Du Bois Mentee, National Institutes of Justice (2017 - present)

□ Syracuse University Graduate Assistantship (Summer 2015 - Present)

□ Second Place in the Best Paper Competition at the 15th EWEPA (2017)

□ Korean Government Overseas Study Fellow (2013-2014)

Service & Others

□ Referee: Journal of Productivity Analysis
Discussant: Economics of Crime and Policing Symposium (Syracuse Univ.)

Outreach: Chung-Ang Univ. (Korea)

Citizenship: Republic of Korea, F-1 visa

Languages & Skills: English (Fluent), Korean (Native) & Matlab, R, Stata, \LaTeX

References

William C. Horrace (Primary)  
Distinguished Professor  
Economics, Syracuse Univ.  
(315) 443-9061  
whorrace@maxwell.syr.edu

Alfonso Flores-Lagunes  
Professor  
Economics, Syracuse Univ.  
(315) 443-9045  
afloresl@maxwell.syr.edu

Yoonseok Lee  
Associate Professor  
Economics, Syracuse Univ.  
(315) 443-9529  
ylee41@maxwell.syr.edu