Essays on Characterizing Inefficiency for Stochastic Frontier Models

Ian Anthony Wright
Syracuse University

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

Part of the Social and Behavioral Sciences Commons

Recommended Citation
Dissertations - ALL. 462.
https://surface.syr.edu/etd/462

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.
Abstract

This dissertation consists of three essays on stochastic frontier models in characterizing inefficiency for a cross section of firms in essays one and three and a set of firms overtime in essay two. The first essay looks at stationary points for several models used in stochastic frontier analysis. The second essay extends the multivariate probability statements of Horrace (2005) to calculate the probability that a firm is any particular efficiency rank. These rank probabilities are used to calculate expected efficiency ranks for each firm. The third and final essay adds spatial correlation to the production function of each firm and generalizes the Horrace (2005) probability statements.

The skew of Ordinary Least Squares (OLS) residuals of the composed error is expected to be negative and positive for a production function and a cost function, respectively. However, because of sampling errors in empirical applications, modelers may get a positive skew for a production function and this has serious implication for Maximum Likelihood Estimates (MLEs) – this is called the “wrong skew problem”. Waldman (1982) shows that for the normal-half normal model if the wrong skew occurs then (1) MLEs reduce to OLS, (2) this solution is stable, and (3) there is a relationship between the skew of OLS residuals and the MLE of the pretruncated variance of inefficiency. In the literature two solutions are provided when the wrong skew occurs; (1) find a new random sample, however this might be too costly and (2) respecify the distribution of inefficiency.

The first essay generalizes part 1 of Waldman (1982) result using the theory of the Dirac measure (Dirac, 1930). This essay shows that if the inefficiency distribution converges to a Dirac delta function when the pretruncated variance of the inefficiency distribution goes to zero, the
likelihood of the composed error will converge to a likelihood based solely on the noise distribution. In particular this essay shows that if the Dirac delta function is centered at zero then the maximum likelihood estimator equals the ordinary least squares estimator in the limit. The parameters of the inefficiency distribution are not identified in the limit. Stability of the maximum likelihood estimator and the “wrong skew” results are derived or simulated for common parametric assumptions on the inefficiency distribution. This essay shows that the full suite of Waldman (1982) result holds for the normal-doubly truncated normal and the normal-truncated normal models when the pretruncated mean is non-positive. Simulation results show that if the wrong skew occurs the MLEs for the normal-doubly truncated normal (when the upper bound $B, B>2\mu$ where $\mu$ is the pretruncated mean), the normal-truncated normal and the normal-exponential models reduce to OLS. A cost function with the wrong skew of OLS residuals is estimated using the Greene’s Airline data and the results show that the normal-truncated normal and the normal-exponential models reduce to OLS. Overall the results reveal that respecifying using the traditional assumptions for the inefficiency distribution is unnecessary if the wrong skew of OLS residuals occurs.

Empirical applications of frontier analysis are abundant ranging from the Airline industry to the farming industry, see Battese and Coelli (1995; 1992), Druska and Horrace (2004) and Almanidis, Qian and Sickles (2014). In empirical applications a modeler typically proceeds by estimating a Cobb-Douglas production function or a cost function for a set of firms. For a production function output is proxy by the total sales deflated by a price index. Inputs include all the factors of production such as land, labor and capital. The first step is to estimate OLS since it provides consistent estimates for all the parameters except the intercept or Corrected Ordinary Least Squares (COLS) which is OLS corrected for the biased intercept. The next step is to
examine the skew (the third central moment) of OLS residuals before proceeding to MLE which is more efficient than OLS. The skew has important information so it is used as a guideline for empiricists as to how to proceed in applications. If the skew has the correct sign (negative for a production function) empiricists proceed to MLE. If the skew has the wrong sign (positive for a production function) empiricists respecify the distribution for inefficiency. This first essay shows that in empirical applications if a modeler encounters the wrong skew respecifying using the normal-truncated or the normal-exponential model is a futile procedure since these models do not provide any new results.

The second essay extends the multivariate probability statements of Horrace (2005) to calculate the conditional probability that a firm is any particular efficiency rank in a sample. Conditional expected efficiency ranks are constructed for each firm, in particular, it can be determined which firm in the sample is the best, $2^{nd}$ best, $\ldots$, $2^{nd}$ worst and worst in the population of firms. Firm level conditional expected efficiency ranks are more informative about the degree of uncertainty in regards to ranking when compared to the traditional ranked efficiency point estimates. A Monte Carlo study reveals that under low skew the expected efficiency rank provides inferential insights which the traditional conditional mean function would not uncover.

The MLEs of the parameters (under the assumption that there are no estimation errors or parameters uncertainty) post estimation are substituted into the conditional mean function. The conditional mean function is the mean of inefficiency conditioned on the composed error and is used to produce estimates for inefficiencies for each firm in the sample, see Jondrow, Lovell, Materov and Schmidt (1982). The probability statements utilize both the first and the second moments which provide a more accurate description of the distribution inefficiency. In empirical
applications to determine which firm is any efficiency rank the modeler substitutes MLEs into
the probability statements (inefficiency conditioned on the composed error) and simulates the probabilities. The firm with the largest probability is interpreted as the best firm in the sample and the firm with the smallest probability is interpreted as the worst firm in the sample. Thereafter the modeler uses these conditional probabilities to compute the expected efficiency rank, such that the firm with the largest value of the expected efficiency rank is deemed the least efficient or ranked the worst and the firm with the smallest value is ranked as the most efficient or the best firm in the sample.

The third essay generalizes the Horrace (2005) probability statements to account for spatial correlation in the unobservable for a cross section of firms. This essay relaxes the assumption of independence on the noise or signal or both noise and signal distributions. This essay makes two assumptions on the inefficiency (signal) distribution, (1) inefficiency is assumed to be truncated from a normal distribution prior to the addition of spatial correlation and (2) inefficiency is drawn from a normal distribution and then truncated. The addition of spatial correlation to the production function results in the likelihood being intractable as the number of integrals increases with the sample size. This essay uses sequential conditioning by Spanos (1986; 1999) to factor the joint distribution into the product of a marginal and univariate conditional distributions to compute the probability of the least and most efficient firm. Unlike Horrace (2005) if inefficiency is assumed to be spatially correlated, the conditional distribution of inefficiency conditioned on the composed error is not needed to compute the probabilities. The MLEs are substituted directly into the probability statements. This is because spatial dependence induces heteroskedascity that results in variation across the firms. Overall this essay
provides some insights to empiricists in making inference when the assumption of independence on the noise and inefficiency distributions is relaxed.

The presence of spatial correlation in the errors shifts the production function outward or inward. The composed error is not random because firms are locating in specific areas due to easier access to specialized workers which reduces the search cost of matching workers to the appropriate firms. Furthermore firms will locate in places where there are more favorable demand conditions, similar cultural practices, bureaucratic organization, work ethics and economic activities. Having better access to inputs will affect the productivity of a given firm, however these activities are not observed by the econometrician. These activities affect efficiency and need to be accounted for empirically to provide a better characterization of inefficiency. The spatial correlation is captured using a prespecified weighted matrix. There are several ways of determining the weights. For instance, a modeler could employ contiguity weights or use an inverse distance function. The inverse distance function means that firms further away from each other will impact each other less. These weights are typically known prior to estimation. The weighted matrix is added to the production function before estimation begins. Post estimation MLEs or COLS are substituted into the probability statements developed in Horrace (2005), in which the modeler will be able to compute the probability that firm $i$ is the least or most efficient firm in the sample.
Essays on Characterizing Inefficiency for Stochastic Frontier Models

Ian A. Wright

B.Sc., The University of the West Indies, 2000

M.Sc., The University of the West Indies, 2002

MA., Virginia Polytechnic and State University, 2006

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate School of Syracuse University.

May 2016
Acknowledgements

I would like to express my deepest gratitude to all those who have helped me to complete this dissertation and contribute to my intellectual growth at Syracuse University.

To my wife, Dr. Sannisha Dale, she has filled my world with warmth and unrelenting support for my career and personal aspirations. Her love and understanding has kept me going throughout this process.

I would like to thank my thesis advisor Professor William Horrace for his guidance and help during the preparation of this dissertation. His ongoing support is priceless and without him this could hardly have been done. I would also like to thank him for sharing his knowledge in helping me to develop and enhance my understanding of econometrics.

My sincere gratitude goes to Professor Chihwa Kao who has helped me and continues to encourage me on this journey. I would also like to thank, Professor Badi Baltagi and Jan Ondrich for their support during my studies.

Also a huge thanks to the econometrics study group in the economics department at Syracuse University, this includes, Bin Peng and Fa Wang. It is an encouraging environment, where econometric ideas are fleshed out and unanswered questions are unearthed. I have also had constructive discussions with Pallab Gosh and Jaeyoon Lee. I would also like to thank the staff of the economics department, in particular Laura Sauta for her continued support on administrative tasks over the past years.

Finally I want to thank my parents, in particular my mother Valrie Moonah who passed away late last year. She believed wholeheartedly in me and encouraged me on this academic journey.
Table of Contents

Essay I: Stationary Points for Parametric Stochastic Frontier Models........... 1

Introduction............................................................................................................. 2
Limiting Behavior of the Likelihood Function.......................................................... 7
Stability of OLS for Common Models................................................................. 13
Simulations.............................................................................................................. 15
Empirical Application........................................................................................... 20
Conclusion............................................................................................................ 22
References ................................................................................................ 23
Appendix ............................................................................................................... 26

Essay II: Expect Efficiency Ranks from Parametric Stochastic Frontier Models.................................................................................................................. 121

Introduction........................................................................................................... 122
Conditional Inefficiency Rank Probabilities For Parametric Frontiers............... 125
Monte Carlo Study................................................................................................. 130
Empirical Example............................................................................................... 135
Conclusion............................................................................................................ 139
References........................................................................................................... 140

Essay III: Probability Statements for Stochastic Frontier Models with Spatial Errors.............................................................................................................. 156

Introduction ........................................................................................................... 157
Violation of the Assumption of Independence on ε............................................. 162
Frontier Models with Spatial Correlation in the Errors........................................ 164
Identification and the Likelihood........................................................................... 170
Likelihood............................................................................................................. 172
Simulations............................................................................................................ 174
The Conditional Distribution............................................................................... 175
Probability of Maximal and Minimal draws of Inefficiency in the presence of Spatial Correlation.......................................................... 177
Conclusion............................................................................................................ 183
References........................................................................................................... 184
Appendix............................................................................................................... 188

List of Tables and Figures

TABLES

ix
Essay II: Expect Efficiency Ranks from Parametric Stochastic Frontier Models

Table 1: Simulation Parameters and the Resulting Truncated Moments ........................ 145
Table 2: Average and Quantiles of Conditional Expected Rank, Low Skew, skew(u) = 0.50 ........................................................................................................ 146
Table 3: Average and Quantiles of Conditional Expected Rank, Medium Skew, skew(u) = 1.0. ........................................................................................................ 147
Table 4: Average and Quantiles of Conditional Expected Rank, High Skew, skew(u) = 1.5. ........................................................................................................ 148

Table 5. Heterogeneous Vessel Efficiency Results Tier 1, Sorted on TE .......... 149
Table 6. Heterogeneous Vessel Efficiency Results Tier 2, Sorted on TE .......... 149
Table 7. Heterogeneous Vessel Efficiency Results Tier 3, Sorted on TE .......... 149

FIGURES

Essay II: Expect Efficiency Ranks from Parametric Stochastic Frontier Models

Figure 1: Low skew: skew(u) = 0.5. ............................................................... 151
Figure 2: Medium skew(u)=1.0... .................................................................... 152
Figure 3: Large skew(u)=1.5. ........................................................................ 153
Figure 4: Kernel Density of Truncated Normal Distribution with low skew: skew(u)=0.5. ................................................................................................. 154
Figure 5: Kernel Density of Truncated Normal Distribution with medium skew: skew (u)=1. ............................................................................................. 154
Figure 6: Kernel Density of Truncated Normal Distribution with medium skew: skew(u)=1.5. ............................................................................................. 155
Essay I: Stationary Points for Parametric Stochastic Frontier Models
1 Introduction

Parametric Stochastic Frontier Analysis (SFA) allows for production function estimation while accounting for inefficiency in a cross-section of firms.\footnote{This paper is concerned with production function estimation, but the analysis can be applied to cost functions as well.} Specifically, SFA models a firm’s output as a function of its inputs plus a random error. Aigner, Lovell and Schmidt (1977), hereafter ALS, specify a composed error $\varepsilon = v - u$, where $v$ represents random fluctuations in the production frontier and where $u \geq 0$ (independent of $v$) represents random inefficiency. Typically $u$ is called signal, $v$ is called noise, and the model is parameterized in terms of a "signal to noise" ratio of the variance components. Estimation proceeds by making distributional assumptions on the error components and calculating (or searching for) the maximum likelihood estimator (MLE). ALS (1977) specify a normal distribution for noise, $v \sim N(0, \sigma_v^2)$, a half normal (HN) distribution for signal, $u \sim |N(0, \sigma_u^2)|$, so that the "Normal-Half Normal" stochastic frontier model (the N-HN model) has signal to noise ratio, $\lambda = \frac{\sigma_u}{\sigma_v}$.\footnote{ALS (1977) also consider an exponential distributional assumption on the inefficiency distribution, leading to a Normal-Exponential model. In the N-HN model the variance of the pretruncated distribution of inefficiency is $\sigma_u^2$. The variance of the post-truncated distribution is $V(u) \neq \sigma_u^2$. The distinction is important in what follows.} The HN specification for $u$ implies that its skew is positive so that skew of $\varepsilon$ is negative. However, in practice it often happens that the skew of the Ordinary Least Squares (OLS) residuals (of a regression of output on inputs) is positive, which implies that the MLE of $\sigma_u$ is equal to zero. This is called the "wrong skew" problem, and all rigorous treatments of the issue in the literature have been for the N-HN specification. See for example, Olson, Schmidt and Waldman (1980), Waldman (1982), Simar and Wilson (2009) and Feng, Horrace and Wu (2013).

Waldman (1982) analyzes the wrong skew problem for the N-HN specification, showing that when the residuals of OLS have the wrong skew: (1) OLS is the MLE (i.e., the MLE of $\sigma_u$ equals zero), (2) OLS is a stable solution in the parameter space of the likelihood, and (3) there is an inverse relationship between the sign of the skew of the OLS residuals and the MLE of $\sigma_u$. The implication for empirical exercises is that inefficiency in the sample is zero for all firms, when the OLS residuals have the wrong skew. Therefore, if a priori it is believed that there is inefficiency in the population of firms, wrong skew of the OLS residuals...
is problematic if the model is estimated using Maximum Likelihood Technique. Theoretically, wrong skew of the OLS residuals creates problems for inference, because the Hessian of the likelihood is singular. This can be overcomed by the bagging technique of Simar and Wilson (2009), however there is still potential for the lack of identification of the models' parameters, as this paper will show. Empirical "solutions" to the "wrong skew" problem include pulling another sample (which is often not practical) or respecifying the distribution of inefficiency. The latter solution begs the question, "do other distributions suffer from the "wrong skew" problem, and if they do what are the implications for identification of the models?" A contribution of this paper is that it considers the issue for other specifications on the distribution of $u$ and shows that respecification maybe redundant.

In his analysis Waldman (1982) exploits the "signal to noise" parameterization of the N-HN model by setting $\sigma_u = 0$, so $\lambda = 0$ in the likelihood. In this case the likelihood is finite at $\lambda = 0$, and it reduces to OLS, making the analysis tractable. A problem with this approach is that when $\sigma_u = 0$ the distribution of inefficiency is pathological. In fact, in the N-HN case, the parameter $\sigma_u$ is restricted to be strictly positive, so setting it equal to zero in the likelihood causes the inefficiency density to be singular. In the N-HN model (with a "signal to noise" parameterization of the likelihood) the singularity is not problematic in determining the behavior of the likelihood in the neighborhood of OLS. This fortuitous outcome does not occur in general, and for many parameterizations of the stochastic frontier model the likelihood is not finite when $\sigma_u$ equals zero. In this regard, this paper develops a general theory that examines the limiting behavior of the likelihood function based on a singular distribution for inefficiency. This is the primary contribution of this paper, and it exploits the "sifting property" of the Dirac measure (Dirac, 1930) to examine models under general distributional assumptions on inefficiency. Most of the common parametric assumptions on inefficiency distributions (e.g., truncated normal or exponential) have a Dirac measure (or Dirac delta) representation when inefficiency variance tends towards zero.\textsuperscript{3} This paper generalizes the "OLS is the MLE"
result of Waldman (1982), and provides sufficient conditions for any parametric specification of the model to achieve this result. In particular this paper shows that as the pretruncated variance of the inefficiency distribution goes to zero, OLS is the MLE as long as $v$ is normally distributed and the distribution of $u$ is a Dirac delta located at the origin. If not, then the location parameter of the Dirac delta function and the intercept of the production function are confounded and are not identified. This may suggest that empiricists restrict their distributional choices for $u$ to classes of functions that have a Dirac delta representation located at the origin. Both the HN and exponential distributions possess this feature, as do the truncated normal (TN) and doubly truncated normal (DTN) distributions when the pre-truncated mean is non-positive. This may be particularly relevant when it is believed that the population of firms under study is marked by low average inefficiency and low variability of inefficiency (i.e., the inefficiency distribution is close to being singular at the origin).

This paper then explores the stability of the OLS solution and the wrong skewness issue for common parameterizations on the distribution of inefficiency by examining the behavior of the likelihood function in the neighborhood of OLS for different models. The following cases are considered: the Normal-Truncated Normal model (N-TN) due to Stevenson (1980), the Normal-Exponential model (N-E) due to ALS (1977) and Meeusen and Van den Broeck (1977), and the Normal-Doubly Truncated Normal model (N-DTN) due to Almanidis, Qian and Sickles (2014). For the N-TN and N-DTN models with a non-positive pretruncated mean of inefficiency, the complete suite of Waldman results are derived: a stable stationary point at OLS where the MLE of $\sigma_u$ equals 0, corresponding to the wrong skewness of OLS residuals. In particular, when the pretruncated mean is non-positive, the behavior of the likelihood in the neighborhood of OLS for these models is identical to that of the N-HN model.

If the pretruncated mean of inefficiency is positive in the N-TN and N-DTN models, then the inefficiency distribution has a Dirac delta representation, but it is centered on a positive location: the minimum of the pretruncated mean and the inefficiency upper bound. In this case neither OLS nor MLE are identified. In
particular the positive pretruncated mean of the inefficiency distribution is not identified (nor is the upper bound of inefficiency for the N-DTN model). This is, perhaps, further evidence that the pretruncated mean of inefficiency is only weakly identified in these models (Almanidis, Qian and Sickles, 2014, p64). The problem is that, as the pretruncated inefficiency variance goes to zero, there is no information to identify the pretruncated mean or the upper bound. Nonetheless, simulations suggest that for the N-TN model the MLE of the variance of inefficiency is indeed zero, when the skew of the OLS residuals is positive.

For the N-E model, stability of OLS is established, but this paper cannot establish a theoretical results on the relationship between the skew of the OLS residuals and the local optimality of OLS. However, simulations suggest that the full suite of Waldman-type results hold. Generally speaking, as the pretruncated variance of the inefficiency distribution tends to zero all of the aforementioned models are observationally equivalent when the location of the mass point of the resulting Dirac delta is zero.

The theoretical results have additional implications for empiricists. As pointed out above for MLE to nest OLS and for it to be identified, empiricists must select distributions for inefficiency that collapse at the origin as variance goes to zero. This is an argument for simpler, single parameter distributions like the HN or the exponential distributions of ALS (1977). Distributional choice is important when faced with wrongly skewed OLS residuals, since a common "solution" to the "wrong skew" problem is to specify a new distribution for inefficiency. Additionally, if the empirical exercises begin with OLS estimation, and the estimation produces wrong skew, then the theoretical results suggest that MLE of the N-TN model is not an option when the pre-truncated mean is non-positive, because it will be observationally equivalent to MLE of the N-HN model. If empiricists are convinced that inefficiency exists and restricts the MLE of the pretruncated mean to be positive then this will not be a feasible solution because the location parameter only shifts the distribution, but the shape of the distribution is preserved. Therefore if the wrong skew occurs respecifying using the N-TN model is not an option even if empiricists are convinced that inefficiency exists in the population. This is also true (but to a lesser extent) for the N-DTN model. If the OLS residuals have positive skew, then
MLE of the N-DTN can accommodate this by estimating a parameterization of the DTN distribution with negative skew (so that $-u$ has positive skew). To accommodate positive skew, the MLE of the pretruncated mean must be positive (and it must be larger than 1/2 of the estimated upper bound). Therefore, when faced with the wrong skew of the OLS residuals and the N-DTN specification is chosen, it makes sense to restrict the pretruncated mean to be positive (and perhaps greater than 1/2 of the upper bound) in the estimation. Unfortunately, if the distribution of inefficiency is truly singular, then neither the N-TN nor the N-DTN models will be identified when the pretruncated mean is positive. These nuances of the empirical implementations of parametric SFA underscore the difficulties of the implicit deconvolution problem that the composed error model generates. Such difficulties are exacerbated when there is only a cross-section of data to aid in estimating the model parameters. These findings suggest when faced with incorrectly skewed OLS residuals empiricists should either admit that inefficiency does not exist in the sample or (if another sample is not available) use the inferential procedures of Simar and Wilson (2010).

The simulated results of this paper take these empirical conclusions a step further. The simulations suggest that even when the pre-truncated mean of the N-TN model is positive, incorrectly skewed residuals imply that OLS is a stable and optimal solution to MLE of the N-TN model. Therefore, when the wrong skew arises, the N-HN and N-TN models are observationally equivalent (regardless of the sign of the pretruncated mean), and neither of the specifications are appropriate if it is believed that inefficiency exists in the population. Unfortunately, simulated evidence also suggests that this may be true for the N-E model, implying that when faced with incorrectly skewed residuals, the N-E model will also not be a "solution" to wrong skew.

This paper is organized as follows. The next section establishes the main result on maximum likelihood estimation of the parametric stochastic frontier model when the distribution of inefficiency is singular with a Dirac delta function representation. Section 3 provides theoretical stability results for the N-E model and

---

4 This is related to doing MLE on the N-HN model, but restricting the pre-truncate variance to be positive. See Feng, Horrace and Wu (2013).

5 See Horrace and Parmeter (2011) for a discussion of deconvolution in cross-sectional SFA.
for the N-TN and N-DTN models when the pretruncated mean is non-positive. In the latter two cases it shows the relationship between the skew of the OLS residuals and the stability of OLS. Section 4 presents simulated results for the N-DTN, N-TN and N-E models. Section 5 provides an empirical application using the Greene’s Airline Data. Section 6 summarizes and concludes.

2 Limiting Behavior of the Likelihood Function

The cross-sectional stochastic frontier model of ALS (1977) is:

\[ y_i = x_i^\prime \beta + \varepsilon_i, \quad i = 1, ..., n \]  

(1)

where \( y_i \) is a single output (typically in logarithms), \( x_i \) is a \( k \times 1 \) vector of inputs with the first element equal to 1 for all \( i \), \( \beta \) is a \( k \times 1 \) vector of unknown parameters, and \( \varepsilon_i = v_i - u_i \) represents random shocks to the production process. The \( v_i \) are random fluctuations in the production frontier for each firm \( i \), and the \( u_i \) are random inefficiency draws for each firm \( i \). Without specific distributional assumptions on the error terms, the basic assumptions of the ALS (1977) model are:

Assumption 1 \( v_i \) and \( u_i \) are independent.

Assumption 2 \( v_i \) and \( u_i \) are independent of \( x_i \).

Assumption 3 \( v_i \in (-\infty, \infty) \) has absolutely continuous probability density \( f_v(v, \sigma_v) \).

Assumption 4 \( u_i \geq 0 \) has absolutely continuous probability density \( f_u \) with variance parameter \( \sigma_u > 0 \).

These are generally accepted assumptions throughout the literature regardless of the specific parametric form of the inefficiency distribution in Assumption 4. The density of \( u \) is known up to \( \sigma_u \) (and perhaps other
parameters that are suppressed for now) and write: \( f_u(u, \sigma_u) \).\(^6\) Given these assumptions, the composed error has continuous density:

\[
f_f(\varepsilon) = \int_0^\infty f_v(\varepsilon + u, \sigma_v) f_u(u, \sigma_u) du.
\]

The likelihood function is:

\[
L(y, x, \beta, \sigma_v, \sigma_u) = \prod_{i=1}^n f_f(y_i - x_i' \beta) = \prod_{i=1}^n \int_0^\infty f_v(y_i - x_i' \beta + u, \sigma_v) f_u(u, \sigma_u) du.
\] (2)

This paper centers on the behavior of the (log) likelihood function at the point \( \sigma_u = 0 \). Waldman (1982) is able to do this in the N-NH model by setting \( \sigma_u = 0 \), observing that the solution to maximizing the likelihood is equivalent to the OLS estimator of \((\beta, \sigma_v)\), and examining the behavior of the Hessian at OLS to determine that it is a stable solution. He also concludes that at \( \sigma_u = 0 \) the skew of the OLS residuals is necessarily positive: the "wrong skew" as compared to the negatively skewed composed error. In general \( f_u(u, 0) \) may not be well-defined, so plugging \( \sigma_u = 0 \) into the likelihood may not always be an option in understanding the behavior of OLS in the parameter space of MLE.\(^7\) Therefore, to understand the likelihood one must consider its behavior as \( \sigma_u \to 0 \). Based on Assumptions 1-4, the likelihood is:

\[
\lim_{\sigma_u \to 0} L(y, x, \beta, \sigma_v, \sigma_u) = \prod_{i=1}^n \lim_{\sigma_u \to 0} f_f(y_i - x_i' \beta) = \prod_{i=1}^n \lim_{\sigma_u \to 0} \int_0^\infty f_v(y_i - x_i' \beta + u, \sigma_v) f_u(u, \sigma_u) du.
\] (3)

To understand the limiting behavior of the likelihood is to understand the limiting behavior of the integral on the RHS of equation 3, which is governed by the limiting behavior of \( f_u \) inside the integral. Therefore,

this paper makes the following additional assumption on this density.\(^8\)

\(^6\)For example, in ALS (1977) the distribution of \( u \) is known up to \( \sigma_u \) and can either be HN, \( u \sim N(0, \sigma_u^2) \), or exponential, \( f_u(u, \sigma_u) = \frac{1}{\sigma_u} e^{-\sigma_u u} \). The N-TN and N-DTN models have additional unknown parameters in the density of \( u \) and the likelihood function. Although the additional parameters make for a much richer class of models, they make estimation more difficult in general. Moreover, these additional parameters are not identified when \( \sigma_u \to 0 \).

\(^7\)This is certainly the case with the N-DTN, N-TN and N-E models. The paper considers the behavior of the likelihood at OLS for all these models in this paper.

\(^8\)Moving from Equation 3 to Equation 4 requires limit and integral to be interchanged as demonstrated below. For probability densities that have lebesgue measure, Dominating convergence theorem (DCT) can be used to allow the interchanging of limit
Assumption 5  As \( \sigma_u \to 0 \), \( f_u(u, \sigma_u) \) approximates a Dirac delta, \( \delta(u - a) \), with mass point at \( a \geq 0 \).

When \( \sigma_u \to 0 \) the continuous cumulative density possesses a "big jump" and becomes discontinuous, defined as \( F(u - a) \). The discontinuous density is \( \delta(u - a) = \frac{d}{du}F(u - a) \). According to Bracewell (2000) \( F(u - a) \) does not possess a derivative at \( u = a \) and we should interpret \( \delta(u - a) \) as "the derivative of the sequence of differentiable functions that approach \( F(u-a) \) as a limit constitute a suitable defining sequence for \( \delta(u-a) \)". According to Griffiths (2005) the Dirac delta function can be thought of as the limit of a sequence of functions of ever-increasing height and ever-decreasing width. The Dirac delta is a symmetric function such that \( \delta(u - a) = 0 \) for \( u \neq a \), and \( \delta(u - a) \to \infty \) for \( u = a \), satisfying the property \( \int_{-\infty}^{\infty} \delta(u - a)du = 1 \). Essentially, it is zero everywhere except for an infinitely large singularity at \( u = a \), yet the area under the curve is still unity. This apparent contradiction arises because the Dirac delta is not a function per se, but serves as the representation of a limiting process which is useful under the Reimann integral.\(^9\) See Frieden (1983) of Arley and Buch (1950) for a measure theoretic definition of the Dirac delta.\(^10\) Most of the usual parametric assumptions on the density of \( u \) satisfy Assumption 5, and in particular Gaussian functions are known to approximate Dirac deltas.

The delta has the following integration properties.

**Lemma 1 (Delta Properties)** Let \( g(u) \) be a finite function that is continuous in a closed neighborhood of

and integral. However, the sequences of probability density converge to a Dirac delta function which has a Dirac measure and this is in conflict with the classical lebesgue measure therefore DCT is not applicable. Cheng (2006), Proposition (2.3) and Theorem 2.4 shows that differentiation and integration can be interchanged for generalized functions (the Dirac delta function is one type of a generalized function). Cheng (2006) provides an example using the Dirac delta function in which integral and differentiation is interchanged. Differentiation is a limit hence it can be deduced that limit and integral can be interchanged. Note that the sequences of functions \( f_{\theta_n}(x, \theta_n) \) are integrable almost everywhere and there exists only a finite number of discontinuities or singularities, which can be ignored, because change in the value of a function at a single point does not affect the integral. The limit of the function \( f_{\theta}(x, \theta) = \delta(x - a) \) is integrable (\( \int_{0}^{\infty} \delta(x - a)dx = 1 < \infty \)) with respect to a dirac measure. The limit of the integral converges to the integral of the limit and the order of limit and integration (whether doing the integration and then taking the limit or taking the limit and then doing the integration) is trivial.

\(^9\)Technically it is a *generalized function* and \( f_u \) can be thought of converging pointwise as such we can write \( \delta(u - a) = \lim_{\sigma_u \to 0} f_u(u, \sigma_u) \).

\(^10\)For \( x \in S \), \( F_x(A) \) is defined on \( S \) such that: \( F_x(A) = \{1, x \in A\} \) where \( F_x \) is a measure on \( S \) see, Zitkovic lecture note.
Then,

\[ \int_{-\infty}^{\infty} g(u)\delta(u-a)du = g(a) \]  \hspace{1cm} (i)

\[ \int_{-\infty}^{\infty} g(u) \frac{d^n}{du^n}\delta(u-a)du = (-1)^n \frac{d^n}{du^n}g(a) \]  \hspace{1cm} (ii)

Property (i) and (ii) are well-known results and can be shown by integration by parts, see Kobayashi (2009). Property (i) is called the sifting property and is useful for understanding the limiting behavior of the integral (over \( u \)) in equation 2. Note that the operation of \( g(u) \) on the left-hand side of property (i) sifts out a single value of \( g(u) \), which is \( a \). This is the importance of \( \delta(u-a) \) because irrespective of the type of functions that are in the integrand, \( \delta(u-a) \) allows for the flexibility to deal with them and only the integral matters. Under Assumptions 1-5, the likelihood is:

\[ \lim_{\sigma_u \to 0} L(y, x, \beta, \sigma_v, \sigma_u) = \prod_{i=1}^{n} f_v(y_i - x_i' \beta + a, \sigma_v). \]  \hspace{1cm} (4)

This follows by directly applying the sifting property of a Dirac delta function. Equation 4 generalizes a result in Waldman (1982). Horrace and Parmeter (forthcoming) attempt to generalize Waldman (1982) result where they assume that \( v \) has a zero-mean Laplace distribution. Their proof uses limiting arguments on the characteristic functions of the error components. There is a subtle distinction between their proof and the one presented here. They rely on the assumption that the limit of the characteristic function of \( u \) equals one as \( \sigma_u \to 0 \). Their unit characteristic function implies that the density of inefficiency must be degenerate in the limit \( (\lim_{\sigma_u \to 0} f_u(u, \sigma_u) = 1 \) for \( u = a = 0 \) otherwise), which is a much stronger assumption than that implied by \( \delta(u-a) \). Also, working with the Dirac delta in the space of the density function of the composed error is more natural when analyzing the behavior of the likelihood.
When $\sigma_u \rightarrow 0$ there is no variability in the draws from the inefficiency distribution. Without this variability none of the parameters of the inefficiency distribution are identified. The Hessian of the likelihood function has at least one eigenvalue equal to zero, and all models are observationally equivalent up to $a$ regardless of the specification of $f_u$. In fact, the lack of variability in the inefficiency distribution causes $a$ to be "not identified" in general. The lack of identification of $a$ induces a lack of identification of the intercept in the production function (the first element of $\beta$). Therefore, the only models worth considering in the limit are those with $a$ (ex ante) equal to 0. The paper also notes that in the limit the usual conditional mean of $u$, $\lim_{\sigma_u \rightarrow 0} E(u|\varepsilon_i)$, is constant across $i$ and is equal to $\lim_{\sigma_u \rightarrow 0} E(u) = a$. That is, the variability in $\varepsilon$ provides no information on the moments of $u$ and the distribution of $u$ collapses to its mean at $a$. A special case of Equation 4 is when $v$ is a zero-mean normal random variable.

**Result 1** For $v \sim N(0, \sigma_v^2)$, MLE is equal to OLS if $a = 0$.

The proof follows by substituting $a = 0$ into $\lim_{\sigma_u \rightarrow 0} L$ in Equation 4, and recognizing that the resulting likelihood is equivalent to that of OLS with normal errors. Result 1 says that in the limit, MLE equals OLS displaced by an unknown constant $a > 0$. This is the classical "lack of identification" of the constant term in OLS when the mean of the errors is a non-zero constant, but it is exacerbated by the fact that positive $a$ cannot be estimated (is not identified). Therefore, empiricists need to select $f_u$ with $a = 0$ for the usual "MLE equals OLS" result to emerge as $\sigma_u \rightarrow 0$. This is unlike the "corrected OLS" estimator of the stochastic frontier literature (Horrace and Schmidt, 1996), where a priori information is needed on $f_u$ to correct the OLS constant.

The idea of requiring $a = 0$ so that $\lim_{\sigma_u \rightarrow 0} f_u(u, \sigma_u) = \delta(u)$, is related to a priori beliefs on the nature of inefficiency as $\sigma_u \rightarrow 0$. The current thinking in the literature is that "$\sigma_u = 0$" is synonymous with "zero inefficiency" in the population of firms, which is equivalent to the density of $u$ collapsing to zero as $\sigma_u \rightarrow 0$. Indeed, this is true for the N-HN model, N-TN and N-DTN (with non-positive pretruncated mean), as its

---

11See Rothenberg (1971) for a discussion on the concept of observational equivalence.
Dirac delta is centered on $a = 0$. However, it is possible for certain specifications of $f_u$ to collapse to positive $a$. In these cases, is it reasonable to think that a population of firms might be "stuck" at some positive levels on inefficiency? This is a philosophical question that will not be addressed here, but it may be worth considering when specifying a distribution for $u$ for parametric SFA, particularly when one may believe that the pretruncated variance of inefficiency is close to zero.

The common distributions are introduced for $u$ that satisfy assumption 5 and this paper calculates the singularity point, $a$, in each case. The most general truncated form of a normal distribution in the literature is the DTN by Almanidis, Qian and Sickles (2014), in which case:

$$f_u(u) = \frac{1}{\sigma_u \phi\left(\frac{u-\mu}{\sigma_u}\right)} \left[ \Phi\left(\frac{B-\mu}{\sigma_u}\right) - \Phi\left(\frac{-\mu}{\sigma_u}\right) \right] ; \ u \in [0, B], \ B > 0$$

where $\phi$ and $\Phi$ are the standard normal pdf and cdf, respectively. In this DTN specification: zero is the lower truncation point, $\mu$ is the pre-truncated mean, and $B > 0$ is the upper truncation point. This nests the TN specification (where $B \to \infty$) and the HN specification (where $B \to \infty$ and $\mu = 0$). This paper also considers the exponential specification where $f_u(u) = \exp\{-u/\sigma_u\}/\sigma_u, u \geq 0$.

**Lemma 2** Common specifications satisfying Assumption 5 are:

i) doubly truncated normal with $a = \max[0, \min(\mu, B)]$.

ii) exponential, with mean $\sigma_u$ and $a = 0$.

The proof is provided in the appendix. Consider the DTN specification of $f_u$. Since $B > 0$, the only time that the condition, $a = 0$, in result 1 will be satisfied is when $\mu \leq 0$. This is also true for the TN (where $B \to \infty$, so $a = \max[0, \mu]$). In the HN model $B \to \infty$ and $\mu = 0$, so $a = \max[0, 0] = 0$, corresponding to the Waldman (1982) result. Also the condition in result 1 ($a = 0$) is always satisfied in the exponential specification. The DTN specification is the only case where the skew of $u$ could be negative and this occurs when $B < 2\mu$. Since $B > 0$ and $\mu > 0$, if the skew of the DTN density is negative, then $a$ is guaranteed to
be positive, and the condition that assures that MLE equals OLS in result 1 is violated.

**Observational Equivalence** From Lemma 2, when $\sigma_u \to 0$ and $\mu \leq 0$, the doubly truncated normal, the truncated normal, the half normal and the exponential specifications are all identical in the limit. The inefficiency parameters are not identified.

Let $\theta^1 = (\sigma^2, \lambda, B) \neq \theta^2 = (\sigma^2, \lambda, \mu) \neq \theta^3 = (\sigma^2, \lambda) \neq \theta^4 = (\sigma_u, \sigma_v)$ be the parameter vector associated with the N-DTN, N-TN, N-HN and N-E models, respectively. When $\sigma_u \to 0$, for $\mu \leq 0$ it implies that $a = 0$, all the above models are identical, that is $\lim_{\sigma_u \to 0} f^dtn(u, \theta^1) = \lim_{\sigma_u \to 0} f^tn(u, \theta^2) = \lim_{\sigma_u \to 0} f^hn(u, \theta^3) = \lim_{\sigma_u \to 0} f^{exp}(u, \theta^4)$, and hence the modeler is unable to discriminate among the different specifications on $u$. The likelihoods are indistinguishable and hence the parameters for the inefficiency distributions are not identified.

When $\sigma_u \to 0$, for $\mu > 0$ it implies that $a \neq 0$, the DTN and TN specifications are observational equivalence if ($\mu < B$). If the wrong skew occurs respecifying using traditional models will not provide any new result.

The paper now considers the stability of the OLS solutions for all of these models. In some cases the paper provides theoretical results for a stable solution, corresponding to OLS residuals having the wrong skew. In other cases there is no theoretical solution forthcoming. In these cases the paper considers simulated evidence in a subsequent section.

### 3 Stability of OLS for Common Models

Waldman-type results are derived for some common specification of the stochastic frontier model.

**Theorem 1** If $\mu \leq 0$, OLS is a stable stationary point in the likelihoods of the normal-doubly truncated normal, the normal-truncated normal, and the normal-half normal specification of the parametric stochastic frontier model. In all cases, the "wrong skewness" of the OLS residuals corresponds to OLS being a local maximum in the parameter space of MLE.
The proof is provided the appendix. Theorem 1 generalizes the result of Waldman (1982) to the N-TN and N-DTN models, but in these cases you cannot follow Waldman (1982) and simply plug $\sigma_u = 0$ into the likelihood. The condition in equation 4 that the pretruncated mean ($\mu$) needs to be non-positive is related to the requirement in result 1 that $a = 0$ in the Dirac delta function. Although it is not a result per se, the calculations suggest that MLE/OLS equivalence, stability of OLS, and the "wrong skew" results hinge critically on $a = 0$ as $\sigma_u \to 0$. That is, the TN and DTN specifications of the density of inefficiency both have $a = 0$ when $\mu \leq 0$, as does the HN specification. Not surprisingly the final results on the behavior of the likelihood in the neighborhood of OLS are similar to Waldman’s result. Walman (1982) calculates the change in the likelihood in the neighborhood of OLS to be:

$$\Delta \ln L(y, x, \beta, \sigma_u, \sigma_u = 0) = \frac{1}{6} \frac{\sum e_i^3}{\sigma_u^3} \frac{2}{\sqrt{2\pi}} \frac{\pi - 4}{\pi} \gamma^3 + o(\gamma^4), \quad (6)$$

where $e_i$ is the OLS residual, $\sigma_u^3 = (\sum e_i^2 / n)^{3/2}$, and $\gamma$ is a small, positive number, representing a perturbation of the likelihood away from OLS. Since $(\pi - 4)/\pi < 0$, $\Delta \ln L$ is the opposite sign of $\sum e_i^3$, the skew of the OLS residuals. If this skew is "correct" (negative), then the likelihood increases ($\Delta \ln L > 0$) as we move away from OLS (i.e., as $\sigma_u$ becomes positive). If the skew is wrong (positive), OLS is a local maximum.

Similarly, the calculations show that:

$$\text{N-DTN: } \lim_{\sigma_u \to 0} \Delta \ln L(y, x, \beta, \sigma_u, \sigma_u, \mu \leq 0, B) = \frac{1}{6} \frac{\sum e_i^3}{\sigma_u^3} \frac{2}{\sqrt{2\pi}} \frac{\pi - 4}{\pi} \gamma^3 + o(\gamma^4), \quad (7)$$

$$\text{N-TN: } \lim_{\sigma_u \to 0} \Delta \ln L(y, x, \beta, \sigma_u, \sigma_u, \mu \leq 0) = \frac{1}{6} \frac{\sum e_i^3}{\sigma_u^3} \frac{2}{\sqrt{2\pi}} \frac{\pi - 4}{\pi} \gamma^3 + o(\gamma^4). \quad (8)$$

An important difference between equation 6 and the new equations 7 and 8 is that the former equation simply substitute $\sigma_u = 0$ into $\Delta \ln L$ while the latter equations require calculation of the limit as $\sigma_u \to 0$ of the $\Delta \ln L$. Therefore, both the N-DTN and N-TN MLEs behave exactly like the N-HN MLE in the
neighborhood of OLS when $\mu \leq 0$, and the complete suite of Waldman (1982) results apply.\footnote{For the N-DTN model: $\text{plim} \left( \frac{1}{n} \Sigma \varepsilon^3 \right) \rightarrow E[\varepsilon - E(\varepsilon)]^3 = -E[u - E(u)]^3 < 0$ for $\sigma_u^2 \neq 0$ and $\mu \leq 0$, and skew is positive if $(B < 2\mu)$, see Almanidis, Qian and Sickle (2014).} For the N-E model this paper shows that:

**Theorem 2** OLS is a stable stationary point in the likelihood of the normal-exponential specification of the parametric stochastic frontier model.

For the N-E model this paper shows that as $\lim_{u \to 0} \triangle LnL(y, x, \beta, \sigma_v, \sigma_u) = 0$. For the N-E model there is no theoretical relationship between the skew of OLS residuals and the MLE of $\sigma_u$, see appendix.

4 Simulations

The simulations are designed to examine the relationship between the skew of the OLS residuals and the MLE of the pretruncated variance parameter $\sigma_u$ in all models considered. In cases where a theoretical relationship is established, the purpose is to verify the results. In cases where no theoretical relationship could be established, the purpose is to see if such a relationship may exist. Simulations are performed for the N-E model, the N-TN model and the N-DTN model when the skew is positive ($B > 2\mu$). In all cases the simulation sample size is 1,000, and various sample sizes ($n = 25, 50, 100$) and signal to noise ratios ($\lambda = \frac{\sigma_u}{\sigma_v} = 0.25, 0.50, 1.0$), with the usual restriction that $\sigma_u^2 + \sigma_v^2 = 1$. This paper varies the pretruncated means ($-0.5, -0.1, 0.1, 0.5$) for the N-TN model and both the pretruncated means ($-0.5, 0.1, 0.5$) and the upper bounds ($1.2, 0.5$) for the N-DTN model. The paper selects relatively small values for the signal to noise ratio and sample size to ensure that there are sufficient cases where the OLS residuals have negative skew. The data generation process (DGP) for both the N-TN and N-DTN models is the simple specification, $y = \varepsilon = v - u$, similar to the simulation study conducted in ALS (1977). The DGP for the N-E model includes a constant $y = 3 + v - u$.\footnote{The simulations are conducted in Matlab 7.4.0 version. The paper uses unrestricted MLE to estimate all three models. The function fminunc is used to maximize function. This uses the BFGS Quasi-Newton method with mixed quadratic and cubic line}

\footnote{For the N-DTN model : $\text{plim} \left( \frac{1}{n} \Sigma \varepsilon^3 \right) \rightarrow E[\varepsilon - E(\varepsilon)]^3 = -E[u - E(u)]^3 < 0$ for $\sigma_u^2 \neq 0$ and $\mu \leq 0$, and skew is positive if $(B < 2\mu)$, see Almanidis, Qian and Sickle (2014).}
This paper attempts to estimate the N-TN and N-DTN models with and without a constant, however these models do not perform well with a constant and it is still unclear why this is the case.\footnote{14} One explanation could be that unlike the N-HN and the N-E models which are globally identified, the N-TN and N-DTN models have local identifications and as such may be more difficult to estimate with a constant.\footnote{15} Additionally, according to Greene (The Econometric Approach to Efficiency Analysis, page 45) estimating a non-zero mean quite frequently impedes or prevents the convergence of the iterations for the N-TN model. Furthermore, Greene (E62 Stochastic Frontier Models and Analysis, page 50) noted that it is difficult to distinguish between the pretruncated mean ($\mu$) and the parameter $\sigma_u$ in this model. That is $\mu$ and $\sigma_u$ can covary so that there is little or no variation in the mean ($E(u)$) of $u$. The likelihood is a function of the pretruncated ($\mu$) so any information regarding $u$ and $\sigma_u$ is informed by the pretruncated mean. When there is little variation in $u$ it increases the difficulty of estimating the pretruncated mean and it might be the case that the intercept could be picking up some of this effect.

The loglikelihoods that are used to estimate all three models are given in equation 9, 10 and 11 for N-DTN, N-TN and N-E models, respectively. The loglikelihoods for the N-DTN and N-TN models have the standard lambda parameterization. For the N-DTN a value for the upperbound of $B = 0$ can be treated as the same as $\sigma_u \to 0$ which means that there is no probability mass for the distribution of $u$. This is beyond the scope of this paper and will not be pursued. In the simulation when this paper attempts to use $B = \mu = 0$ for starting values for the N-DTN, the loglikelihood yields a log value of 0 which is negative infinity and it poses problems numerically. The model fails to converge and does not provide an estimate for any of the parameters of the model. The loglikelihood of the N-TN model does not suffer from this problem when $\mu = 0$, see equation 10. The N-E likelihood has a scale factor of $\frac{1}{\sigma_u}$, hence whenever the wrong skew occurs the estimate of the pretruncated variance of inefficiency is set close to zero which implies that the loglikelihood will have a starting value of positive infinity, see equation 11. However, in the simulations search; it uses the BFGS formula for updating the approximation for the Hessian.

\footnote{14}This paper simulates these models with one regressor and their performances are satisfactory. 
\footnote{15}See Rothenberg (1971) for a discussion on the distinction between local and global identification.
this does not pose a problem numerically for the loglikelihood of the N-E model. The N-HN estimates are used for the starting values for both the N-TN and N-DTN models, except for B where this paper uses the maximum value of the residuals for each draw. For the N-E model, OLS estimates are used for the starting values.

\[ \text{N-DTN: } \ln l(\beta, \sigma^2, \lambda, \mu, B; y) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^5 (y - x' \beta + \mu)^2}{2\sigma^2} - n \ln[\Phi\left( \frac{(B - \mu)(\lambda^2 + 1)^{\frac{1}{2}}}{\sigma} \right) - \Phi\left( \frac{-\mu(\lambda^2 + 1)^{\frac{1}{2}}}{\sigma} \right)] + \sum_i^n \ln[\Phi\left( \frac{(B + (y - x' \beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma} \right) - \Phi\left( \frac{(y - x' \beta)\lambda - \mu\lambda^{-1}}{\sigma} \right)] \] (9)

\[ \text{N-TN: } \ln l(\beta, \sigma^2, \lambda, \mu, y) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^5 (y - x' \beta + \mu)^2}{2\sigma^2} - n \ln[1 - \Phi\left( \frac{-\mu(\lambda^2 + 1)^{\frac{1}{2}}}{\sigma} \right)] + \sum_i^n \ln[1 - \Phi\left( \frac{(y - x' \beta)\lambda - \mu\lambda^{-1}}{\sigma} \right)] \] (10)

\[ \text{N-E: } \ln l(\beta, \sigma_v, \sigma_u, y) = -n \ln(\sigma_u) + \sum_i^n \ln[1 - \Phi\left( \frac{(y - x' \beta)}{\sigma_v} + \frac{\sigma_v}{\sigma_u} \right)] + \sum_i^n \left( \frac{(y - x' \beta)}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right) \] (11)

This paper will discuss the N-DTN model, then the N-TN model and finally the N-E model. For the N-DTN model the results are in figures 1 to 12. The signal to noise ratios \( \lambda = 0.25, 0.50, 1.0 \) correspond to different values for the parameter, \( \sigma_u = 0.2425, 0.4472, 0.7071 \), respectively. Each figure contains three panels (a, b, and c) corresponding to each of the three sample sizes, \( n = 25, 50, 100 \), respectively. In the panels each circle represents one of 1,000 simulations draws, where the MLE of \( \sigma_u \) and the values of the skew of OLS residuals are recorded. This paper first fixes \( \mu \) and \( B \) and varies \( \lambda \). For figure 1, panel a, for \( \mu = -0.5, B = 1.2 \) for a sample size of 25, 52.0% of the sample has the wrong (positive) skew of OLS residuals for \( B > 2\mu \) which is a noisy experiment. As the sample increases to \( n = 50 \) and \( n = 100 \), panel b and c respectively, the proportions of wrong (positive) skew decline to 50.1% and 49.4%, respectively. As one moves across panels (a to c), the cloud of estimates with the correct (negative) skew become more dense as \( n \) increases. The results are the same in figures 2 and 3 except the signal to noise ratio increases \( (\lambda) \) to 0.5.
and 1.0, respectively, for fixed values of $\mu$ and $B$. The proportion of wrong (positive) skew of OLS residuals decreases to 41.3 % for $n = 100$ corresponding to $\lambda = 1.0$ and $\sigma_u = 0.7071$. For figures 4 to 12, either $B$ is fixed and $\mu$ is varied or $B$ is varied and $\mu$ is fixed for the same $\lambda$ values and the conclusion is the same. For example in figure 8, for $\mu = -0.5$ and $B = 0.5$ panel (a-c) for $n = 25, \lambda = 0.25$ the proportion of wrong (positive) skew is 50.4% while for a sample size $n = 100$ the proportion reduces to 47.8%.

Figures 13 to 23 for panel (a-c) show the simulation results for the N-TN model. This paper first fixes $\mu$, for figure 15 panel c, $\lambda = 1.0$, $\mu = -0.5$ and $n=25$ and it shows that the proportion of wrong skew of OLS residuals is 47.0% but as $n$ increases the proportion of wrong (positive) skew decreases to 34.3% for $n = 100$. The cloud also becomes more dense as $n$ increases and this is true across all figures and across all panels (a-c). For the N-TN the simulations results suggest that there is a relationship between skew of OLS residuals and a positive $\mu$. For figure 18, where $\lambda = 0.25$, $\mu = 0.1$, $n = 25$, panel a 47.9% of the sample has the wrong (positive) skew which indicates that the MLE of $\sigma_u$ is zero. The proportion of wrong skew decreases to 45.8% for $n = 100$. Figures 13 to 23 show that there is a negative relationship between the skew of OLS residuals and the MLE of the pretruncated variance of inefficiency irrespective of the sign of the pretruncated mean.

The N-TN and N-DTN models appear to perform better with a relatively small pretruncated mean. A relatively large mean ($\mu$) for the N-TN and N-DTN models resemble the normal density which results in the skew being zero and this might cause an identification problem for $\mu$. For $\mu = 0.5$ both models perform poorly, that is the loglikelihood does not converge and the estimates appear to be deterministic, however as the pretruncated mean reduces there are improvements in both models. The simulation results confirm the theoretical results derived whenever the wrong (positive) skew occurs the MLE of $\sigma_u$ is zero. Also the parameters of the inefficiency distribution $\mu$ and $\sigma_u$ for the N-TN model and $\mu$, $\sigma_u$ and $B$ for the N-DTN model) are not identified when the wrong (positive) skew occurs. Simulation results also confirm that the

\textsuperscript{16}The N-E and N-HN models generalize OLS and is a special case when $\sigma_u \rightarrow 0$, that is skew is zero if and only if $\sigma_u \rightarrow 0$. This is different from the N-TN and N-DTN models which generalize OLS, but is not a special case when $\sigma_u \rightarrow 0$, since the skew can be zero if $\mu$ becomes relatively large for fix value of $\sigma_u$. 

18
\[ E(u) = 0 \text{ for } \mu \leq 0 \text{ and } E(u) = \mu > 0 \text{ when } \sigma_u \to 0. \] A final note is that if the skew of OLS residuals turns out to be positive (wrong) then respecification using the N-TN model and restricting \( \mu > 0 \), with the hope of getting better estimates will be a fruitless exercise because \( \mu \) is a location parameter and this will only shift the distribution and the shape of the distribution will remain unchange.

For the N-E model in figures 24 to 26, for \( \lambda = 0.25, 0.50, 1.0, \) respectively, correspond to different values for the pretruncated inefficiency variance, \( \sigma_u = 0.2425, 0.4472, 0.7071, \) respectively. For figure 24, panel a, a value of \( \lambda = 0.25 \), and a sample size of \( n = 25 \) is a noisy experiment and there are many instances of wrong (positive) skew of the OLS residuals. In fact 33.7\% of the 1,000 simulation draws possess the wrong (positive) skew and for all these draws the MLE of \( \sigma_u \) equals to zero. Moving to panels b and c in figure 24 the sample sizes of \( n = 50 \) and \( 100 \), respectively, it is shown that the frequency of experiments with the wrong (positive) skew decrease to 27.5\% and to 18.3\%, respectively. As can also be seen the cloud of estimates with correct (negative) skew become more dense as \( n \) increases. The results are similar in figures 25 and 26 except the signal to noise ratios have increased to \( \lambda = 0.50 \) and to 1.0, respectively, across the figures, so the frequency of "wrong skew" draws is declining across figures for a given panel. For example, the frequency of wrong (positive) skew draws is declining across each panel a from 33.7\% to 25.2\% to 12.1\%, as one moves from figure 24 to figure 26. In all cases the MLE of \( \sigma_u \) equals zero whenever the skew of the OLS residuals turns out to positive. Therefore, while a theoretical relationship between the skew of OLS residuals and MLE estimate of the pretruncated variance cannot be established, the simulation results suggest otherwise.

For all three models the simulation results show that there is a relationship between the wrong skew of OLS residuals and the MLE of \( \sigma_u \), see figures 1 to 26 for the N-TN, N-DTN and N-E models, respectively. Whenever OLS skew is positive (wrong) the MLE of \( \sigma_u \) is zero. As \( \sigma_u \to 0 \), the N-DTN for \( \mu \leq 0 \), the N-TN for \( \mu \leq 0 \), the N-HN and the N-E models are observational equivalence. Thus, for empiricists facing incorrectly skewed OLS residuals respecifying using either the N-E Model or N-TN model is not an option because simulations suggest that if empiricists select the N-E model the MLE will be equivalent
to OLS, resulting in zero inefficiency in the sample. Empiricists must also be mindful of the fact that an estimate of zero for $\sigma_u$ does not mean that inefficiency in the sample is necessarily zero, it just means that Maximum Likelihood cannot provide an estimate for the true pretruncated variance of inefficiency. Therefore in empirical application if we have a random sample in which empiricists are convinced that inefficiency exists and if the wrong skew occurs, they will have to resort to an alternative methodology to provide inference about $u$.

5 Empirical Application

This section estimates the Greene Airline data for a cost function.\textsuperscript{17} It is a panel dataset which consists of six firms over fifteen years from 1970-1984. This paper ignores the panel structure and provides estimates for the pooled OLS, the N-TN model and the N-E model. The variables are total cost, output (which measures the revenue of passenger miles), fuel price and load factor. The skew of OLS residuals is negative (wrong skew) for a cost function. The cost function to be estimated is:

$$\ln Cost_i = \alpha_0 + \alpha_1 \ln output_i + \alpha_2 \ln fuel_i + \alpha_3 load_i + \varepsilon_i$$  \hspace{1cm} (12)

where $\varepsilon_i = v_i - u_i$

The estimates are in Table 1.

\textsuperscript{17}The data are available on Greene’s New York University website at: http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm
||Table 1: Cost function of Airline Data (1970-1984)||

<table>
<thead>
<tr>
<th></th>
<th>Ordinary Least Square</th>
<th>Normal-Exponential</th>
<th>Normal-Truncated Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>constant</strong></td>
<td>9.517 *</td>
<td>9.517</td>
<td>9.517</td>
</tr>
<tr>
<td></td>
<td>(0.2292)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td><strong>ln output</strong></td>
<td>0.883*</td>
<td>0.883</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>(0.0133)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td><strong>ln fuel</strong></td>
<td>0.454*</td>
<td>0.454</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td>(0.0203)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td><strong>load</strong></td>
<td>-1.623*</td>
<td>-1.623</td>
<td>-1.623</td>
</tr>
<tr>
<td></td>
<td>(0.3453)</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td><strong>σ_v</strong></td>
<td>0.1246</td>
<td>0.1246</td>
<td>0.1246</td>
</tr>
<tr>
<td><strong>σ_u</strong></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>μ</strong></td>
<td>–</td>
<td>–</td>
<td>not identified</td>
</tr>
</tbody>
</table>

*significant at the 5% level.

In the empirical application the skew of OLS residuals is used as a guideline on how to proceed in estimating the frontier. I first estimate the pooled OLS (column 1) and the skew of the OLS residuals is negative (wrong) for the cost function, therefore according to Waldman (1982) the N-HN model is not applicable. I respecify and estimate the cost function using the N-E and N-TN models and the MLE for both models are reduced to OLS. The Hessian is singular and the standard errors cannot be computed.\(^{18}\) This empirical result reveals that in the presence of the wrong skew of OLS residuals both the N-E and N-TN MLEs are reduced to OLS, which implies that respecifying using N-E model or N-TN model does not provide any new results.\(^{19}\)

\(^{18}\) According to Simar and Wilson (2010) conventional standard errors estimates for MLE are unavailable due to singularity of the Hessian.

\(^{19}\) All the models are estimated in Matlab. Additionally, I estimated the frontier in STATA 12.0. The estimates for the
6 Conclusion

This paper shows that the signal distribution is very important in determining the behavior of the likelihood of the composed error when the pretruncated variance goes to zero. It uses Dirac delta theory to show that for any given $u$ and $v$ assume in SFA when $\sigma_u \to 0$, the composed error distribution is determined only by information regarding the noise distribution, if $u$ converges to a Dirac delta function in the limit.

Waldman (1982) full suite holds for the N-TN and N-DTN models for a non-positive pretruncated mean. This paper is unable to find a theoretical result between the skew of OLS residuals and MLE estimate of the pretruncated variance of inefficiency for the N-E model, however simulations suggest otherwise. The inefficiency parameters are not identified when the wrong skew occurs for the respective models.

If the wrong skew occurs, respecifying using the traditional models will not provide any new result. In empirical application if the wrong skew occurs and empiricists are using the N-TN model, then respecifying using the N-TN and restricting the pretruncated mean to be strictly positive will be a fruitless exercise since this will only shift the distribution and the shape will remain the same. Empiricists may respecify using the N-DTN model since it accommodates a negative skew, however the pretruncated mean should be restricted to be positive.

Future research should explore under which conditions a priori information can be used to determine if the pretruncated mean is positive or negative before the modeler proceeds to use Maximum Likelihood technique to estimate the frontier. However, this might only be beneficial for the N-DTN model. If the N-TN model is initially assumed and prior information suggests that the pretruncated mean is positive, and OLS residuals have the wrong skew then use the N-DTN model and restrict the mean to be strictly positive. However, if priori information suggests that the pretruncated mean is negative then it will be a futile exercise to respecify using N-DTN model when the wrong skew occurs. Future research might also look into the behavior of MLEs for the N-DTN model if the true inefficiency distribution has negative skew technology parameters are the same. STATA provides an estimate for the pretruncated mean, but it is insignificant. The N-TN model is reduced to OLS, however the estimates have slightly different standards errors. The N-E model does not converge.
and the skew of OLS turns out to be negative.

Overall this paper has provided some insights and guidelines to empiricists when estimating stochastic frontier models. If the wrong skew of OLS residuals occurs respecifying requires more thought than previous literature suggests.

References


Appendix: Proofs of Lemma and Theorems

This appendix provides all the proofs for the lemma and theorems in the text.

A Proof of Lemma 2

Proof.

The limiting behavior of the doubly truncated normal density is governed by the limiting behaviors of the numerator and the denominator in equation 5. We consider three cases. First, if $\mu \in (0, B)$, then the
limit of the denominator:
\[
\lim_{\sigma_u \to 0} \left\{ \Phi \left( \frac{B - \mu}{\sigma_u} \right) - \Phi \left( \frac{-\mu}{\sigma_u} \right) \right\}
\]

(13)

is a finite and positive constant. Therefore, the limiting behavior of the DTN is dictated solely by limiting behavior of the numerator in equation 5:
\[
\lim_{\sigma_u \to 0} \frac{1}{\sigma_u} \phi \left( \frac{u - \mu}{\sigma_u} \right).
\]

(14)

Since the numerator (divided by the finite and positive limit of the denominator) is proportional to the density of a \( N(\mu, \sigma_u^2) \) random variate with \( \mu \in (0, B) \), then the limit of the DTN density in this case is a Dirac delta with mass point at \( a = \mu \).

The two remaining cases to consider are \( \mu \leq 0 \) and \( \mu \geq B \). In both these cases the limits of the denominator in 13 and the numerator in 14 equal zero. Taking derivatives of these expressions with respect to \( \sigma_u \), and applying L’Hopital’s rule yields:
\[
\lim_{\sigma_u \to 0} f_u(u) = \lim_{\sigma_u \to 0} \frac{1}{\sigma_u} \phi \left( \frac{u - \mu}{\sigma_u} \right) - \frac{\mu}{\sigma_u} \phi \left( \frac{u - \mu}{\sigma_u} \right) - \frac{B - \mu}{\sigma_u} \phi \left( \frac{B - \mu}{\sigma_u} \right).
\]

(15)

Some algebra on equation 15 yields:
\[
\lim_{\sigma_u \to 0} f_u(u) = \lim_{\sigma_u \to 0} \frac{1}{\mu} \exp \left( -\frac{u(u - 2\mu)}{2\sigma_u^2} \right) \left[ 1 - \frac{(u - \mu)^2}{\sigma_u^2} \right].
\]

For \( \mu \leq 0 \) the limit of the denominator above equals 1, so we need only evaluate the limit of the numerator. That is:
\[
\lim_{\sigma_u \to 0} \frac{1}{\mu} \exp \left( -\frac{u(u - 2\mu)}{2\sigma_u^2} \right) \left[ 1 - \frac{(u - \mu)^2}{\sigma_u^2} \right].
\]

In general the limit of the exponential term dominates both the \( \frac{1}{\mu} \) term and the limit of the bracketed term.
For $\mu \leq 0$ the limit of the exponential term is 0, except for $u = 0$ when it equals 1. When $u = 0$ two things may occur. First, if $\mu < 0$, then $\frac{1}{\mu}$ is a negative constant and the bracketed term goes to negative infinity in the limit, so the numerator goes to positive infinity in the limit. Second, if $\mu = 0$, then $\frac{1}{\mu} \to \infty$ and the bracketed term equals 1, so (again) the numerator goes to positive infinity in the limit. Therefore, when $\mu \leq 0$, $\lim_{\sigma_u \to 0} f_u$ is a Dirac delta centered on $a = 0$.

For the $\mu \geq B$ case algebra on equation 15, yields:

$$\lim_{\sigma_u \to 0} f_u(u) = \lim_{\sigma_u \to 0} \frac{\frac{1}{\mu-B} \exp \left(-\frac{u(u-2\mu)-B(B-2\mu)}{2\sigma_u^2}\right)}{-\frac{1}{\mu-B} \exp \left(\frac{B(B-2\mu)}{2\sigma_u^2}\right) + 1}.$$  

For $\mu \geq B$ the limit of the denominator above equals 1, so again we need only evaluate the limit of the numerator. That is:

$$\lim_{\sigma_u \to 0} \frac{1}{\mu-B} \exp \left(-\frac{u(u-2\mu)-B(B-2\mu)}{2\sigma_u^2}\right) \left[1 + \frac{(u-\mu)^2}{\sigma_u^2}\right].$$  

Again, in general the limit of the exponential term dominates the $\frac{1}{\mu-B}$ term and the limit of the bracketed term. For $\mu \geq B$ (and noting that $u \leq B$) the limit of the exponential term is 0, except for $u = B$ when it equals 1. When $u = B$ two thing can occur. First, if $\mu > B$, then $\frac{1}{\mu-B}$ is a positive constant and the bracketed term goes to positive infinity in the limit, so the numerator goes to positive infinity in the limit. Second, if $\mu = B$, then $\frac{1}{\mu-B} \to \infty$ and the bracketed term equals 1, so (again) the numerator goes to positive infinity in the limit. Therefore, when $\mu \geq B$ it must be true that $\lim_{\sigma_u \to 0} f_u$ is a Dirac delta centered on $a = B$. Thus, all three cases are

$$a = \begin{cases} 
0 & \mu \leq 0 \\
\mu & \mu \in (0,B) \\
B & \mu \geq B
\end{cases},$$

and the lemma is proved for the DTN case.
For the exponential density we have: \( \lim_{\sigma_u \to 0} f_u(u) = \lim_{\sigma_u \to 0} \exp\{-u/\sigma_u\}/\sigma_u \), which equals zero for \( u > 0 \), but goes to positive infinity for \( u = 0 \), so it must be true that \( \lim_{\sigma_u \to 0} f_u \) is a Dirac delta centered on \( a = 0 \), and proof of the lemma is complete.

B Figures
Figure 1. Normal-Doubly Truncated Model, $\lambda = 0.25$, $\mu = -0.5$, $B = 1.2$, $\sigma_u = 0.2425$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>52.0%</td>
</tr>
<tr>
<td>50</td>
<td>50.1%</td>
</tr>
<tr>
<td>100</td>
<td>49.4%</td>
</tr>
</tbody>
</table>
Figure 2. Normal-Doubly Truncated Model, $\lambda = 0.5$, $\mu = -0.5$, $B = 1.2$, $\sigma_{\alpha} = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>49.6%</td>
</tr>
<tr>
<td>50</td>
<td>49.3%</td>
</tr>
<tr>
<td>100</td>
<td>48.0%</td>
</tr>
</tbody>
</table>
Figure 3. Normal-Doubly Truncated Model, $\lambda = 1.0$, $\mu = -0.5$, $B = 1.2$, $\sigma_a = 0.7071$, $n = 25, 50, 100$.

(a) Frequency of wrong skew

(b) Frequency of wrong skew

(c) Frequency of wrong skew

<table>
<thead>
<tr>
<th>n</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>45.3%</td>
</tr>
<tr>
<td>50</td>
<td>44.1%</td>
</tr>
<tr>
<td>100</td>
<td>41.3%</td>
</tr>
</tbody>
</table>
Figure 4. Normal-Doubly Truncated Model, $\lambda = 0.25$, $\mu = 0.5$, $B = 1.2$, $\sigma_u = 0.2425$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>48.8%</td>
</tr>
<tr>
<td>50</td>
<td>46.6%</td>
</tr>
<tr>
<td>100</td>
<td>45.9%</td>
</tr>
</tbody>
</table>
Figure 5. Normal-Doubly Truncated Model, $\lambda = 0.5$, $\mu = 0.5$, $B = 1.2$, $\sigma_u = 0.4472$, $n = 25, 50, 100$.

(a)  
(b)  
(c)

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>47.8%</td>
</tr>
<tr>
<td>50</td>
<td>46.0%</td>
</tr>
<tr>
<td>100</td>
<td>45.4%</td>
</tr>
</tbody>
</table>
Figure 6. Normal-Doubly Truncated Model, $\lambda = 1.0$, $\mu = 0.5$, $B = 1.2$, $\sigma_u = 0.7071$, $n = 25, 50, 100$. 

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>47.35%</td>
</tr>
<tr>
<td>50</td>
<td>46.2%</td>
</tr>
<tr>
<td>100</td>
<td>45.3%</td>
</tr>
</tbody>
</table>
Figure 7. Normal-Doubly Truncated Model, $\lambda = 0.25, \mu = -0.5, B = 0.5, \sigma_u = 0.2425, n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>52.6%</td>
</tr>
<tr>
<td>50</td>
<td>50.5%</td>
</tr>
<tr>
<td>100</td>
<td>48.9%</td>
</tr>
</tbody>
</table>
Figure 8. Normal-Doubly Truncated Model, $\lambda = 0.5$, $\mu = -0.5$, $B = 0.5$, $\sigma_a = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50.4%</td>
</tr>
<tr>
<td>50</td>
<td>50.2%</td>
</tr>
<tr>
<td>100</td>
<td>47.8%</td>
</tr>
</tbody>
</table>
Figure 9. Normal-Doubly Truncated Model, $\lambda = 1.0$, $\mu = -0.5$, $B = 0.5$, $\sigma_a = 0.7071$, $n = 25, 50, 100$.

(a)  
(b)  
(c)  

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>49.5%</td>
</tr>
<tr>
<td>50</td>
<td>48.1%</td>
</tr>
<tr>
<td>100</td>
<td>44.9%</td>
</tr>
</tbody>
</table>
Figure 10. Normal-Doubly Truncated Model, \( \lambda = 0.25, \mu = 0.1, B = 0.5, \sigma_u = 0.2425, n = 25, 50, 100 \).

(a) Frequency of wrong skew

\( n \) \( Freq. \)

<table>
<thead>
<tr>
<th></th>
<th>25</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50.7%</td>
<td>50</td>
<td>46.7%</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>48.7%</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>48.7%</td>
<td>48.7%</td>
<td></td>
</tr>
</tbody>
</table>
Figure 11. Normal-Doubly Truncated Model, $\lambda = 0.5, \mu = 0.1, B = 0.5, \sigma_u = 0.4472$, $n = 25, 50, 100$. 

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50.8%</td>
</tr>
<tr>
<td>50</td>
<td>48.4%</td>
</tr>
<tr>
<td>100</td>
<td>47.9%</td>
</tr>
</tbody>
</table>
Figure 12. Normal-Doubly Truncated Model, $\lambda = 1.0$, $\mu = 0.1$, $B = 0.5$, $\sigma_u = 0.7071$, $n = 25, 50, 100$.

Frequency of wrong skew

\begin{aligned}
 n & & \text{Freq.} \\
 25 & & 48.5\% \\
 50 & & 47.3\% \\
 100 & & 46.7\%
\end{aligned}
Figure 13. Normal-Truncated Model, $\lambda = 0.25$, $\mu = -0.5$, $\sigma_a = 0.2425$, $n = 25, 50, 100$.

(a)

(b)

(c)

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>49.5%</td>
</tr>
<tr>
<td>50</td>
<td>48.2%</td>
</tr>
<tr>
<td>100</td>
<td>47.2%</td>
</tr>
</tbody>
</table>
Figure 14. Normal-Truncated Model, $\lambda = 0.5$, $\mu = -0.5$, $\sigma_n = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>49.4%</td>
</tr>
<tr>
<td>50</td>
<td>47.7%</td>
</tr>
<tr>
<td>100</td>
<td>44.5%</td>
</tr>
</tbody>
</table>
Figure 15. Normal-Truncated Model, $\lambda = 1.0$, $\mu = -0.5$, $\sigma_u = 0.7071$, $n = 25, 50, 100$.

- (a) Frequency of wrong skew for $n = 25$:
  - $F_{0.0}$: 47.0%
  - $F_{0.5}$: 38.3%
  - $F_{1.0}$: 34.3%

- (b) Frequency of wrong skew for $n = 50$:
  - $F_{0.0}$: 47.0%
  - $F_{0.5}$: 38.3%
  - $F_{1.0}$: 34.3%

- (c) Frequency of wrong skew for $n = 100$:
  - $F_{0.0}$: 47.0%
  - $F_{0.5}$: 38.3%
  - $F_{1.0}$: 34.3%
Figure 15. Normal-Truncated Model, $\lambda = 0.25$, $\mu = -0.1$, $\sigma_u = 0.2425$, $n = 25, 50, 100$.

(a)  

(b)  

(c)  

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>53.2%</td>
</tr>
<tr>
<td>50</td>
<td>52.1%</td>
</tr>
<tr>
<td>100</td>
<td>50.4%</td>
</tr>
</tbody>
</table>
Figure 16. Normal-Truncated Model, $\lambda = 0.5$, $\mu = -0.1$, $\sigma_u = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>46.5%</td>
</tr>
<tr>
<td>50</td>
<td>46.0%</td>
</tr>
<tr>
<td>100</td>
<td>45.7%</td>
</tr>
</tbody>
</table>
Figure 17. Normal-Truncated Model, $\lambda = 1.0, \mu = -0.1, \sigma_u = 0.7071, n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>39.4%</td>
</tr>
<tr>
<td>50</td>
<td>37.5%</td>
</tr>
<tr>
<td>100</td>
<td>31.6%</td>
</tr>
</tbody>
</table>
Figure 18. Normal-Truncated Model, \( \lambda = 0.25, \mu = 0.1, \sigma_u = 0.2425, n = 25, 50, 100 \).

\[ (a) \quad (b) \quad (c) \]

Frequency of wrong skew

\[
\begin{array}{ccc}
\text{n} & \text{Freq.} \\
25 & 47.9\% \\
50 & 46.8\% \\
100 & 45.8\%
\end{array}
\]
Figure 19. Normal-Truncated Model, $\lambda = 0.5$, $\mu = 0.1$, $\sigma_n = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>47.6%</td>
</tr>
<tr>
<td>50</td>
<td>45.2%</td>
</tr>
<tr>
<td>100</td>
<td>44.8%</td>
</tr>
</tbody>
</table>
Figure 20. Normal-Truncated Model, $\lambda = 1.0$, $\mu = 0.1$, $\sigma_a = 0.7071$, $n = 25, 50, 100$.

(a) $n = 25$

(b) $n = 50$

(c) $n = 100$

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>39.4%</td>
</tr>
<tr>
<td>50</td>
<td>33.6%</td>
</tr>
<tr>
<td>100</td>
<td>29.9%</td>
</tr>
</tbody>
</table>
Figure 21. Normal-Truncated Model, $\lambda = 0.25$, $\mu = 0.5$, $\sigma_u = 0.2425$, $n = 25, 50, 100$.

Frequency of wrong skew

\begin{tabular}{|c|c|}
\hline
$n$ & $Freq.$ \\
\hline
25 & 50.5\% \\
50 & 49.0\% \\
100 & 46.8\% \\
\hline
\end{tabular}
Figure 22. Normal-Truncated Model, $\lambda = 0.5$, $\mu = 0.5$, $\sigma_u = 0.4472$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>47.9%</td>
</tr>
<tr>
<td>50</td>
<td>46.4%</td>
</tr>
<tr>
<td>100</td>
<td>44.8%</td>
</tr>
</tbody>
</table>
Figure 23. Normal-Truncated Model, $\lambda = 1.0$, $\mu = 0.5$, $\sigma_u = 0.7071$, $n = 25, 50, 100$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>40.7%</td>
</tr>
<tr>
<td>50</td>
<td>31.7%</td>
</tr>
<tr>
<td>100</td>
<td>27.0%</td>
</tr>
</tbody>
</table>
Figure 24. Normal Exponential Model, $\lambda = 0.25$, $\sigma_u = 0.2425$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>$n$</th>
<th>$Freq.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>33.7%</td>
</tr>
<tr>
<td>50</td>
<td>27.5%</td>
</tr>
<tr>
<td>100</td>
<td>18.3%</td>
</tr>
</tbody>
</table>
Figure 25. Normal Exponential Model, $\lambda = 0.5, \sigma_n = 0.4472, n = 25, 50, 100.$
Figure 26. Normal Exponential Model, $\lambda = 1.0$, $\sigma_n = 0.7071$, $n = 25, 50, 100$.

Frequency of wrong skew

<table>
<thead>
<tr>
<th>n</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>12.9%</td>
</tr>
<tr>
<td>50</td>
<td>3.7%</td>
</tr>
<tr>
<td>100</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

C Proof of Theorem 1

The N-DTN Model
The loglikelihood for \( y_i = x_i'\beta + \varepsilon_i \) (\( i = 1, ..., n \)) for IID random variables:

\[
\ln l(\beta, \sigma^2, \lambda, \mu, B; y) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y_i - x_i'\beta + \mu)^2}{2\sigma^2} - n \ln \left[ \Phi \left( \frac{(B - \mu)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \right) \right] \\
- \Phi \left( \frac{-\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \right) + \sum_i^n \ln \left[ \Phi \left( \frac{(B + (y_i - x_i'\beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma} \right) - \Phi \left( \frac{(y_i - x_i'\beta)\lambda - \mu\lambda^{-1}}{\sigma} \right) \right]
\]

where:

\[
\lambda = \frac{\sigma_u}{\sigma_v} \\
\sigma^2 = \sigma_u^2 + \sigma_v^2 \\
\frac{1}{\sigma_u} = \frac{(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma}
\]

Notations that are used for convenience, let:

\[
A_1 = \frac{(B - \mu)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \\
A_2 = -\frac{\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \\
A_3 = \frac{(B + (y_i - x_i'\beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma} \\
A_4 = \frac{(y_i - x_i'\beta)\lambda - \mu\lambda^{-1}}{\sigma}
\]

\[
\ln l(\beta, \sigma^2, \lambda, \mu, B; y) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y_i - x_i'\beta + \mu)^2}{2\sigma^2} \\
- n \ln [\Phi(A_1) - \Phi(A_2)] + \sum_i^n \ln \left[ \Phi(A_3) - \Phi(A_4) \right]
\]
Some facts that are used when $\mu \leq 0$ and $\lambda \to 0$:

Proof.

Fact 1 and Fact 2 are needed in the proof.

Showing Fact 1:

$$
\frac{\phi(A_1)}{\phi(A_2)} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{A_1^2}{2}\right) = \exp\left(\frac{1}{2}[-A_1^2 + A_2^2]\right)
$$

$$
A_1^2 = \frac{(B^2 - 2B\mu + \mu^2)(\lambda^{-2} + 1)}{\sigma^2}
$$

$$
A_2^2 = \frac{\mu^2(\lambda^{-2} + 1)}{\sigma^2} \Rightarrow \frac{1}{2}[-A_1^2 + A_2^2] = \left[-\frac{B(B - 2\mu)(\lambda^{-2} + 1)}{\sigma^2}\right]
$$

$$
\frac{\phi(A_1)}{\phi(A_2)} = \exp\left[-\frac{B(B - 2\mu)^2(\lambda^{-2} + 1)}{2\sigma^2}\right] \to 0 \text{ as } \lambda \to 0 \text{ for } \mu \leq 0
$$

(18)

Showing Fact 2:
\[
\frac{\phi(A_3)}{\phi(A_4)} = \exp(\frac{1}{2}[-A_3^2 + A_4^2]) \\
A_3^2 = \left( \frac{(B + \varepsilon)\lambda + (B - \mu)\lambda^{-1}}{\sigma} \right)^2 \\
A_4^2 = \left( \frac{\varepsilon\lambda - \mu\lambda^{-1}}{\sigma} \right)^2 = \left( \frac{\varepsilon^2\lambda^2 - 2\varepsilon\mu + \mu^2\lambda^{-2}}{\sigma^2} \right) \\
A_3^2 = \left( \frac{(B + \varepsilon)\lambda + (B - \mu)\lambda^{-1}}{\sigma} \right)^2 = \left( \frac{\varepsilon^2\lambda^2 - 2\varepsilon\mu + \mu^2\lambda^{-2}}{\sigma^2} \right) \\
\frac{\phi(A_3)}{\phi(A_4)} = \exp\left[-\frac{B^2\lambda^2 + 2\varepsilon B\lambda^2 + B(B - 2\mu)\lambda^{-2} + 2B(B - \mu) + 2\varepsilon B}{\lambda^2\sigma^2}\right] \to 0 
\]

**(C.0.1 FOC: N-DTN, N-TN and N-HN Models)**

This is an application of Result 1. **FOCs** for all three models when \(\lambda \to 0\). For \(\mu \leq 0\) when \(\lambda \to 0\) the Dirac Delta function informs us that the loglikelihoods for the N-DTN and N-TN models are not a function of \(((\mu, \beta)\text{ and } (\mu))\), respectively.

**N-DTN, **FOCs:**

\[
\frac{\partial \ln l}{\partial \beta} = \frac{\sum_i^n x(y - x'\beta + \mu)}{\sigma^2} - \frac{\lambda \sum_i^n [\phi_i(A_3) - \phi_i(A_4)]x_i}{\sigma [\Phi(A_3) - \Phi(A_4)]} 
\]

\[
\frac{\partial \ln l}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{\sum_i^n (y - x'\beta + \mu)^2}{2\sigma^4} - \frac{1}{2\sigma^4} \frac{n[\phi(A_1)(-(B - \mu)(\lambda^{-2} + 1)^\frac{1}{2})] - \phi(A_2)\mu(\lambda^{-2} + 1)^\frac{1}{2}]}{[\Phi(A_1) - \Phi(A_2)]} \\
+ \frac{1}{2\sigma^3} \frac{\sum_i^n [\phi_i(A_3)(-(B + (y - x'\beta))\lambda + (B - \mu)\lambda^{-1})) - \phi_i(A_4)(-(y - x'\beta)\lambda - \mu\lambda^{-1}))]}{[\Phi(A_3) - \Phi(A_4)]} 
\]
\[
\frac{\partial \ln l}{\partial \lambda} = -\frac{n[\phi(A_1)(B-\mu)(\lambda^{-2} + 1)^{-\frac{1}{2}} - \phi(A_2)(\lambda^{-3}) + \phi(A_2)(-\lambda^{-3} + 1)^{-\frac{1}{2}}]}{[\Phi(A_1) - \Phi(A_2)]} \\
+ \frac{\sum_i \phi_i(A_3)((y - x'\beta + \mu \lambda^{-2}) - \phi_i(A_4))}{[\Phi(A_3) - \Phi(A_4)]}
\]

(22)

\[
\frac{\partial \ln l}{\partial \mu} = -\frac{\sum_i \phi_i(A_3)(-\lambda^{-1})}{\sigma^2} - \frac{n[\phi(A_1)(-\lambda^{-2} + 1)^{\frac{1}{2}} - \phi(A_2)(-\lambda^{-2} + 1)^{\frac{1}{2}}]}{[\Phi(A_1) - \Phi(A_2)]} \\
+ \frac{\sum_i \phi_i(A_3)(-\lambda^{-1}) - \phi_i(A_4)(-\lambda^{-1})}{[\Phi_i(A_3) - \Phi_i(A_4)]}
\]

(23)

\[
\frac{\partial \ln l}{\partial B} = -\frac{n[\phi(A_1)(-\lambda^{-2} + 1)^{\frac{1}{2}}]}{\sigma} + \frac{\sum_i \phi_i(A_3)(\lambda + \lambda^{-1})}{[\Phi_i(A_3) - \Phi_i(A_4)]}
\]

(24)

As \(\sigma_u \to 0 \Rightarrow \lambda \to 0 \Rightarrow A_i \text{ for } (i = 1, \ldots, 4) \to \infty \Rightarrow \phi_i(A_i) = 0 \text{ and } \Phi(A_i) = 1 \text{ for } \mu \leq 0.

For \(\frac{\partial \ln l}{\partial \beta} : \)

\[
\lim_{\lambda \to 0} \frac{\partial \ln l}{\partial \beta} = \frac{\sum_i x_i^*(y - x'\beta + \mu)}{\sigma^2} + \frac{\sum_i A_i^* \sum_i [\phi_i(A_3) - \phi_i(A_4)] x_i}{[\Phi_i(A_3) - \Phi_i(A_4)]}
\]

where \(A_i^* = \frac{-\lambda}{\sigma} \text{ when } \lambda \to 0 \Rightarrow [\frac{A_i^* \sum_i [\phi_i(A_3) - \phi_i(A_4)] x_i}{[\Phi_i(A_3) - \Phi_i(A_4)]}] = 0 = 0
\]

Let:

\[
N_{\beta} = \lim_{\lambda \to 0} [\frac{A_i^* \sum_i [\phi_i(A_3) - \phi_i(A_4)] x_i}{[\Phi_i(A_3) - \Phi_i(A_4)]}]
\]

= 0

60
we apply L’Hopital Rule and note that:

\[ \varepsilon_i = y_i - x_i' \beta \]

\[ \frac{\partial A_4}{\partial \lambda} = \frac{\varepsilon + \mu \lambda^2}{\sigma} = \frac{\varepsilon \lambda^2 + \mu}{\lambda^2 \sigma} \Rightarrow \frac{1}{\lambda^2} = 0 \, , \, \frac{\partial A_3}{\partial \lambda} = -\frac{1}{\sigma} \]

\[ \Rightarrow \frac{\partial A_3}{\partial \lambda} = 0 \text{ as } \lambda \to 0 \]

\[ -A_3 A_4 = -\left( \frac{-\lambda}{\sigma} \right) \left( \frac{\varepsilon \lambda - \mu \lambda^{-1}}{\sigma} \right) \]

\[ \Rightarrow -A_3 A_4 = \left( \frac{\varepsilon \lambda^2 - \mu}{\sigma^2} \right) = -\frac{\mu}{\sigma^2} \]

After some algebra the indeterminate expression above is:

\[
N_\beta = \frac{\exp \left( \frac{-B^2 \lambda^2 + 2 \varepsilon B \lambda + B(B - 2\mu) \lambda^{-2} + 2B(B - \mu) + 2B^2 B}{\sigma^2} \right) \left( -\frac{A_3 A_4}{\sigma^2} \frac{\partial A_4}{\partial \lambda} \right) - \frac{\mu}{\sigma^2} + \exp \left( \frac{-B^2 \lambda^2 + 2 \varepsilon B \lambda + B(B - 2\mu) \lambda^{-2} + 2B(B - \mu) + 2B^2 B}{\sigma^2} \right) \frac{\partial A_3}{\partial \lambda} \left( x_i \right)}{\left( \exp \left( \frac{-B^2 \lambda^2 + 2 \varepsilon B \lambda + B(B - 2\mu) \lambda^{-2} + 2B(B - \mu) + 2B^2 B}{\sigma^2} \right) \left( -\frac{A_3 A_4}{\sigma^2} \frac{\partial A_4}{\partial \lambda} \right) + \exp \left( \frac{-B^2 \lambda^2 + 2 \varepsilon B \lambda + B(B - 2\mu) \lambda^{-2} + 2B(B - \mu) + 2B^2 B}{\sigma^2} \right) \frac{\partial A_3}{\partial \lambda} \right) + 1}\]

Note that the exponential dominates every other terms (using fact 2) and it goes to zero and the denominator equals to 1, so when \( \lambda \to 0 \)

\[ \Rightarrow N_\beta = \left( -\frac{\mu}{\sigma^2} \right) x_i \]
\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \beta} \right] = \frac{\sum x_i(y_i - x_i' \beta + \mu)}{\sigma^2} - \frac{\mu \sum x_i}{\sigma^2}
\]

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \sigma^2} \right] = \hat{\beta}_{mle} = \hat{\beta}_{ols} = (\sum x_i x_i')^{-1} \sum x_i y_i = (X'X)^{-1}(X'Y)
\]

(25)

For \( \frac{\partial \ln l}{\partial \sigma^2} \):

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \sigma^2} \right] = -\frac{n}{2\sigma^2} + \frac{\sum (y_i - x_i' \beta + \mu)^2}{2\sigma^4} - \frac{1}{2\sigma^3} \lim_{\lambda \to 0} \frac{\phi(A_1)\left(-((B - \mu)(\lambda^{-2} + 1)\hat{\lambda})\right) - \phi(A_2)\mu(\lambda^{-2} + 1)\hat{\lambda}}{[\Phi(A_1) - \Phi(A_2)]}
\]

\[
+ \frac{1}{2\sigma^3} \lim_{\lambda \to 0} \frac{\sum \phi_i(A_3)\left(-((B + (y - x_i' \beta))\lambda + (B - \mu)\lambda^{-1})\right) - \phi_i(A_4)\left(-((y - x_i' \beta)\lambda - \mu\lambda^{-1})\right)}{[\Phi(A_3) - \Phi(A_4)]}
\]

Let:

\[
N_{1,\sigma^2} = -\frac{1}{2\sigma^3} \lim_{\lambda \to 0} \frac{\phi(A_1)\left(-((B - \mu)(\lambda^{-2} + 1)\hat{\lambda})\right) - \phi(A_2)\mu(\lambda^{-2} + 1)\hat{\lambda}}{[\Phi(A_1) - \Phi(A_2)]} = 0
\]

\[
N_{2,\sigma^2} = \frac{1}{2\sigma^3} \lim_{\lambda \to 0} \frac{\sum \phi_i(A_3)\left(-((B + (y - x_i' \beta))\lambda + (B - \mu)\lambda^{-1})\right) - \phi_i(A_4)\left(-((y - x_i' \beta)\lambda - \mu\lambda^{-1})\right)}{[\Phi(A_3) - \Phi(A_4)]} = 0
\]

Let \( A_1 = \frac{(-B - \mu)(\lambda^{-2} + 1)\hat{\lambda}}{2\sigma^2} \), let \( A_2 = \frac{\mu(\lambda^{-2} + 1)\hat{\lambda}}{2\sigma^2} \), let \( A_3 = \frac{-(B + \mu)(\lambda^{-2} + 1)\hat{\lambda}}{2\sigma^2} \), and \( A_4 = \frac{-(\lambda^{-2} + 1)\hat{\lambda}}{2\sigma^2} \), after some algebra:
\[ N_{1,s^2} = -n\left(\frac{\exp\left[-\frac{B(B-2\mu)^2}{2\sigma^2}(\lambda^{-2}+1)\right]}{\left(\frac{\lambda^{-2}}{\lambda^{-1}}\right)}\right) \]

\[ = -n\left(\frac{\exp\left[-\frac{B(B-2\mu)^2}{2\sigma^2}(\lambda^{-2}+1)\right]}{\left(\frac{\lambda^{-2}}{\lambda^{-1}}\right)}\right) \]

\[ \exp\left[-\frac{B^2\lambda^2+2\mu B \lambda^2+B(B-2\mu)\lambda^{-2}+2B(B-\mu)+2\sigma B}{\sigma^2}\right] \left(-A_3^* A_3 \frac{\partial A_3}{\partial \lambda}\right) + \]

\[ N_{2,s^2} = \sum_i^n \left(\frac{\exp\left[-\frac{B(B-2\mu)^2}{2\sigma^2}(\lambda^{-2}+1)\right]}{\left(\frac{\lambda^{-2}}{\lambda^{-1}}\right)}\right) \]

\[ \exp\left[-\frac{B^2\lambda^2+2\mu B \lambda^2+B(B-2\mu)\lambda^{-2}+2B(B-\mu)+2\sigma B}{\sigma^2}\right] \left(-A_3^* A_3 \frac{\partial A_3}{\partial \lambda}\right) \]

Note that: \(-A_2 A_2^* = -\left[ -\frac{\sigma(\lambda^{-2}+1)}{\mu(\lambda^{-2}+1)} \right] = \frac{\mu^2\lambda^{-2}}{\sigma^2} + \frac{\mu^2}{\sigma^2} \) and \( \frac{\partial A_2}{\partial \lambda} = -\frac{\mu^2}{\sigma^2} \)

\[ \Rightarrow -A_2 A_2^* + \frac{\partial A_2}{\partial \lambda} = \frac{\mu^2\lambda^{-2}}{\sigma^2} + \frac{\mu^2}{\sigma^2} - \frac{1}{2\sigma^2} \]

\[ \Rightarrow N_{1,s^2} = \frac{\mu^2\lambda^{-2}}{2\sigma^4} + \frac{\mu^2}{2\sigma^4} - \frac{1}{2\sigma^2} \]

the denominator in \(N_{1,s^2}\) expression is 1

Note that: \(-A_4 A_4^* = \frac{\epsilon^2\lambda^2-2\epsilon\mu+\mu^2}{2\sigma^4} \) and \( \frac{\partial A_4}{\partial \lambda} = \frac{-\epsilon+\mu \lambda^{-2}}{\epsilon+\mu \lambda^{-2}} = -\frac{1}{2\sigma^2} \)

\[ \Rightarrow -A_4 A_4^* + \frac{\partial A_4}{\partial \lambda} = \frac{\epsilon^2\lambda^2-2\epsilon\mu+\mu^2}{2\sigma^4} - \frac{1}{2\sigma^2} \]

since the exponential term dominates it therefore means that all the terms equal to zero that is multiplied by it.
\[ N_2\sigma^2 = \frac{\varepsilon^2 \lambda^2 - 2\varepsilon \mu + \mu^2 \lambda^{-2}}{2\sigma^4} - \frac{1}{2\sigma^2} \]

the denominator in \( N_2\sigma^2 \) expression is \(-1\)

\[ \Rightarrow \lim_{\lambda \to 0} \left[ -n(N_1\sigma^2) + \Sigma_i^n(N_2\sigma^2) \right] = -\frac{n\mu^2}{2\sigma^4} - \frac{\mu \Sigma_i^n \varepsilon_i}{\sigma^4} \]

substitute \( \lim_{\lambda \to 0} \left[ -n(N_2\sigma^2) + \Sigma_i^n(N_2\sigma^2) \right] \) into \( \frac{\partial \ln l}{\partial \sigma^2} \):

\[
\begin{align*}
\lim_{\lambda \to 0} \frac{\partial \ln l}{\sigma^2} &= \frac{-n}{2\sigma^2} + \frac{\Sigma_i^n(y-x'\beta)^2 + 2\Sigma_i^n \varepsilon_i \mu + n \mu^2}{2\sigma^4} - \frac{n\mu^2}{2\sigma^4} - \frac{\Sigma_i^n \varepsilon_i}{\sigma^4} \\
\lim_{\lambda \to 0} \frac{\partial \ln l}{\partial \sigma^2} &= \frac{-n}{2\sigma^2} + \frac{\Sigma_i^n(y-x'\beta)^2}{2\sigma^4} \Rightarrow \hat{\sigma}_{MLE}^2 = \hat{\sigma}_{OLS}^2 \tag{26}
\end{align*}
\]

For \( \frac{\partial \ln l}{\partial \lambda} \):

\[
\begin{align*}
\lim_{\lambda \to 0} \frac{\partial \ln l}{\partial \lambda} &= \lim_{\lambda \to 0} \frac{n(\phi(A_1)(B-\mu)/(\lambda-2+1)^{-\frac{1}{2}}(-\lambda^{-3}) - (\phi(A_2)\frac{\mu}{\sigma}(\lambda^{-2}+1)-\frac{1}{2}(\lambda^{-3})))}{\Phi(A_1) - \Phi(A_2)} + \\
&\quad \frac{\Sigma_i^n(\phi_i(A_3)(B+(y-x'\beta)-(B-\mu)\lambda^{-2} - \phi_i(A_4)((y-x'\beta)+\mu \lambda^{-2}))}{\Phi(A_3) - \Phi(A_4)} = 0
\end{align*}
\]

\[
\begin{align*}
N_1\lambda &= \lim_{\lambda \to 0} \frac{n(\phi(A_1)(B-\mu)/(\lambda-2+1)^{-\frac{1}{2}}(-\lambda^{-3}) - (\phi(A_2)\frac{\mu}{\sigma}(\lambda^{-2}+1)-\frac{1}{2}(\lambda^{-3})))}{\Phi(A_1) - \Phi(A_2)} = 0 \\
N_2\lambda &= \lim_{\lambda \to 0} \frac{\Sigma_i^n(\phi_i(A_3)(B+(y-x'\beta)-(B-\mu)\lambda^{-2} - \phi_i(A_4)((y-x'\beta)+\mu \lambda^{-2}))}{\Phi(A_3) - \Phi(A_4)} = 0
\end{align*}
\]

Let \( A_1^* = -(B-\mu)/(\lambda^{-2}+1)^{-\frac{1}{2}}(\lambda^{-3}) \), \( A_2^* = \mu/(\lambda^{-2}+1)^{-\frac{1}{2}}(\lambda^{-3}) \), \( A_3^* = (B+x)-(B-\mu)\lambda^{-2} \) and \( A_4^* = \epsilon + \mu \lambda^{-2} \)

After some algebra:
\[ N_1 = -n \left( -A_1^* A_1 \exp\left[ -\frac{B(B-2\mu)^2}{2\sigma^2}(\lambda^{-2}+1) \right] \left( \frac{\partial A_1^*}{\partial \lambda} + \frac{\partial A_1}{\partial \lambda} \right) + A_2^* A_2 - \frac{\partial A_2^*}{\partial \lambda} \right) \]

\[ N_2 = \sum_i^n \left( \exp\left[ -\frac{B_2^2 \lambda^4 + 2cB^2(\lambda^2+1) + 2B_2(B-\mu)\lambda^2 + 2cB^2}{\lambda^2\sigma^2} \right] \left( \frac{\partial A_3^*}{\partial \lambda} + \frac{\partial A_3}{\partial \lambda} \right) + A_4^* A_4 - \frac{\partial A_4^*}{\partial \lambda} \right) \]

Note that: \(-A_2^* A_2 = -\left[ \left( \frac{\mu}{\sigma^2} \right)^2 \right] \left( \frac{\mu(\lambda^{-2}+1)}{\sigma} \right) = -\mu^2/\lambda^3 \) and \( \frac{\partial A_2^*}{\partial \lambda} = -3/\lambda + 1/\lambda(\lambda^2+1) \)

\[ \Rightarrow -A_2^* A_2 + \frac{\partial A_2^*}{\partial \lambda} = \left[ \frac{\mu^2}{\sigma^2\lambda^3} + \frac{-3}{\lambda} + \frac{1}{\lambda(\lambda^2+1)} \right] \]

Note that: \(-A_4^* A_4 = -\left[ \frac{\lambda^2}{\sigma^2} \right] \left( \frac{\lambda^2+1}{\sigma} \right) = -\frac{\mu}{\sigma^2} \) and \( \frac{\partial A_4^*}{\partial \lambda} = \frac{2\mu}{\lambda(\lambda^2+\mu)} \)

\[ \Rightarrow -A_4^* A_4 + \frac{\partial A_4^*}{\partial \lambda} = \left[ -\frac{\mu^2}{\sigma^2\lambda^3} - \frac{2\mu}{\sigma^2\lambda^3} - \frac{2\mu}{\lambda(\lambda^2+\mu)} \right] \]

since the exponential term dominates, all the terms multiply by it equals to zero in the limit. The denominator equals to \(-1\).

\[ \Rightarrow N_1 = -n \left[ \frac{\mu^2}{\sigma^2\lambda^3} + \frac{-3}{\lambda} + \frac{1}{\lambda(\lambda^2+1)} \right] \]

\[ \Rightarrow N_2 = \sum_i^n \left[ -\frac{\varepsilon^2}{\sigma^2} + \frac{\mu^2}{\sigma^2\lambda^3} - \frac{2\mu}{\lambda(\varepsilon^2+\mu)} \right] \]

\[ \Rightarrow \lim_{\lambda \to 0} N_1 + N_2 = \frac{\mu^2 - \varepsilon^2 \lambda^4 - \mu^2}{\sigma^2\lambda^3} + \frac{-2\mu(\lambda^2+1) + 3(\varepsilon^2+\mu)(\lambda^2+1) - ((\varepsilon^2+\mu))}{\lambda(\varepsilon^2+\mu)(\lambda^2+1)} \]

\[ \Rightarrow \lim_{\lambda \to 0} N_1 + N_2 = \frac{-\varepsilon^2 \lambda + (3\varepsilon^2 + 2\varepsilon\lambda + \mu\lambda)}{(\varepsilon^2+\mu)(\lambda^2+1)} = 0 \]

\[ \Rightarrow \lim_{\lambda \to 0} \left( \frac{\partial \ln l}{\partial \lambda} \right) = 0 \] (27)
For \( \frac{\partial \ln l}{\partial \mu} \):

\[
\lim_{\lambda \to 0} \frac{\partial \ln l}{\partial \mu} = -\frac{n(y-x'\beta + \mu)}{\sigma^2} - \lim_{\lambda \to 0} \left( \frac{n}{\sigma} \frac{\phi(A_1)(-\lambda^2 + 1)^{1/2} - \phi(A_2)(-\lambda^2 + 1)^{1/2}}{[\Phi(A_1) - \Phi(A_2)]} + \sum_i \phi_i(A_3)(-\frac{\lambda - 1}{\sigma}) - \phi_i(A_4)(-\frac{\lambda - 1}{\sigma}) \right) \frac{\Phi_i(A_3) - \Phi_i(A_4)}{[\Phi_i(A_3) - \Phi_i(A_4)]}
\]

Let:

\[
N_1 = \lim_{\lambda \to 0} \left[ -\frac{n}{\sigma} \frac{\phi(A_1)(-\lambda^2 + 1)^{1/2} - \phi(A_2)(-\lambda^2 + 1)^{1/2}}{\Phi(A_1) - \Phi(A_2)} \right] = 0
\]

\[
N_2 = \lim_{\lambda \to 0} \left( \sum_i \phi_i(A_3)(-\frac{\lambda - 1}{\sigma}) - \phi_i(A_4)(-\frac{\lambda - 1}{\sigma}) \right) \frac{\Phi_i(A_3) - \Phi_i(A_4)}{[\Phi_i(A_3) - \Phi_i(A_4)]} = 0
\]

Let \( A_1^* = A_2^* = -\frac{(\lambda^2 + 1)^{1/2}}{\sigma} \) and \( A_3^* = A_4^* = \frac{-\lambda - 1}{\sigma} \)

After some algebra:

\[
N_1 = \left[ \left( \exp \left( \frac{-B(B - 2\mu)^2(1-\lambda^2)}{2\sigma^2} \right) \left( -A_1^* A_3 \frac{\partial A_1}{\partial \lambda} + \frac{\partial A_1^*}{\partial \lambda} + A_2 A_4^* - \frac{\partial A_4^*}{\partial \lambda} \right) \left( \exp \left( \frac{-B(B - 2\mu)^2(1-\lambda^2)}{2\sigma^2} \right) - 1 \right) \right) \right]
\]

\[
N_2 = \sum_i \left( \exp \left( \frac{-B^2 \lambda^4 + 2B \lambda^2 + B^2 - 2B \lambda^2 + 2B \lambda^2}{\lambda^2 \sigma^2} \right) \right) \left( -A_3^* A_4^* \frac{\partial A_3}{\partial \lambda} + \frac{\partial A_3^*}{\partial \lambda} + A_4 A_4^* - \frac{\partial A_4^*}{\partial \lambda} \right) \left( \exp \left( \frac{-B^2 \lambda^4 + 2B \lambda^2 + B^2 - 2B \lambda^2 + 2B \lambda^2}{\lambda^2 \sigma^2} \right) - 1 \right)
\]

Note \(-A_2 A_2^* = \left[ \frac{-\mu(\lambda^2 + 1)^{1/2}}{\sigma} \right] \frac{-(\lambda^2 + 1)^{1/2}}{\sigma} = -\mu(\lambda^2 + 1) + \frac{1}{\mu} \Rightarrow -A_2 A_2^* + \frac{\partial A_2}{\partial \lambda} = \mu(\lambda^2 + 1) + \frac{1}{\mu} \)

Note \(-A_4 A_4^* = \left[ \frac{\lambda - \mu \lambda - 1}{\sigma} \right] \left( \frac{\lambda - 1}{\sigma} \right) = \frac{-\mu(\lambda - 1)}{\sigma} \) and \( \frac{\partial A_4}{\partial \lambda} = \frac{\lambda - 2}{\sigma^2 + \mu} \)
\[ \Rightarrow -A_4 A^*_1 + \frac{\partial A_1^*}{\partial \lambda} = \frac{\varepsilon - \mu \lambda^{-2}}{\sigma^2} + \frac{1}{\varepsilon \lambda^2 + \mu} \]

the exponential terms dominates, all the terms equal to zero which is multiplied by it when \( \lambda \to 0 \). The denominator equals to \(-1\)

\[
N1_{\mu} = \lim_{\lambda \to 0} [-n(\frac{-\mu(\lambda^{-2} + 1)}{\sigma^2} + \frac{1}{\mu})]
\]

\[
N2_{\mu} = \lim_{\lambda \to 0} \sum_i^n \left( \frac{\varepsilon - \mu \lambda^{-2}}{\sigma^2} + \frac{1}{\varepsilon \lambda^2 + \mu} \right)
\]

\[
\lim_{\lambda \to 0} [-N1_{\mu} + N2_{\mu}] = -[\frac{-\mu(\lambda^{-2} + 1)}{\sigma^2} + \frac{1}{\mu}] + \frac{\varepsilon - \mu \lambda^{-2}}{\sigma^2} + \frac{1}{\varepsilon \lambda^2 + \mu} = \frac{\mu + \mu \lambda^2 - \mu}{\sigma^2} + \frac{\varepsilon}{\sigma^2} - \frac{1}{\mu} + \frac{1}{\varepsilon \lambda^2 + \mu} = \frac{\mu}{\sigma^2} + \frac{\varepsilon}{\sigma^2}
\]

\[
\Rightarrow \lim_{\lambda \to 0} \left( \frac{\partial \ln l}{\partial \mu} \right) = -\frac{\sum_i^n (y_i - \bar{x} i \beta)}{\sigma^2} - \frac{n \mu}{\sigma^2} + \frac{n \mu}{\sigma^2} + \frac{\sum_i^n (y_i - \bar{x} i \beta)}{\sigma^2} = 0 \quad \text{(28)}
\]

For \( \frac{\partial \ln l}{\partial B} \):

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial B} \right] = -\frac{n}{\sigma} \lim_{\lambda \to 0} \left[ \frac{\phi(A_1)(-(\lambda^{-2} + 1)\frac{1}{2})}{[\Phi(A_1) - \Phi(A_2)]} + \frac{\sum_i^n \phi_i(A_3)(\frac{\lambda + \lambda^{-1}}{\sigma})}{[\Phi_i(A_3) - \Phi_i(A_4)]} \right] = 0
\]

\[
N1_B = -\frac{n}{\sigma} \lim_{\lambda \to 0} \left[ \frac{\phi(A_1)(-(\lambda^{-2} + 1)\frac{1}{2})}{[\Phi(A_1) - \Phi(A_2)]} \right]
\]

\[
N2_B = \lim_{\lambda \to 0} \left[ \frac{\sum_i^n \phi_i(A_3)(\frac{\lambda + \lambda^{-1}}{\sigma})}{[\Phi_i(A_3) - \Phi_i(A_4)]} \right]
\]

Let \( A_1^* = \frac{(\lambda^{-2} + 1)\frac{1}{2}}{\sigma} \) and \( A_3^* = \frac{\lambda + \lambda^{-1}}{\sigma} \)

After some algebra:
Since the exponential term dominates, all the terms equal to zero in the numerator for both $N_1B$ and $N_2B$ and the denominator equals to $-1$ in both terms.

\[
\lim_{\lambda \to 0} \left[ -\frac{n}{\sigma} \left( \exp\left[ -\frac{B(B-2\mu)^2}{2\sigma^2} \right] (-A_1^* A_1 \frac{\partial A_1}{\partial A}) \right) \right] = 0
\]

\[
\lim_{\lambda \to 0} \sum_{i} \left[ \exp\left[ -\frac{B^2 \lambda^4 + 2e B \lambda^4 + B(B-2\mu) + 2B(B-\mu)\lambda^2 + 2e B \lambda^2}{\lambda^2 \sigma^2} \right] (-A_3^* A_3 \frac{\partial A_3}{\partial A}) \right] = 0
\]

Since the exponential term dominates, all the terms equal to zero in the numerator for both $N_1B$ and $N_2B$ and the denominator equals to $-1$ in both terms.

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial B} \right] = 0
\] (29)

Note that $\sigma = \sigma_v$ after we substitute $\sigma_u \to 0$.

The $\ln l(\beta, \sigma^2, \lambda, \mu, B; y)$ for $N-DTN$:

the MLEs=OLS in the limit:

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \beta} \right] = \hat{\beta}_{ols}
\] (30)

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \sigma^2} \right] = \tilde{\sigma}_v^2
\] (31)

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \lambda} \right] = 0
\] (32)

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \mu} \right] = 0
\] (33)
\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial B} \right] = 0
\]  
(34)

The \( \ln l(\beta, \sigma^2, \lambda; \mu; y) \) for N-TN:

first \( B \to \infty \) and then allow \( \lambda \to 0 \),

the MLEs=OLS in the limits:

\[
\lim_{\lambda \to 0} \lim_{B \to \infty} \left[ \frac{\partial \ln l}{\partial \beta} \right] = \hat{\beta}_{ols}
\]  
(35)

\[
\lim_{\lambda \to 0} \lim_{B \to \infty} \left[ \frac{\partial \ln l}{\partial \sigma^2} \right] = \sigma_v^2
\]  
(36)

\[
\lim_{\lambda \to 0} \lim_{B \to \infty} \left[ \frac{\partial \ln l}{\partial \lambda} \right] = 0
\]  
(37)

\[
\lim_{\lambda \to 0} \lim_{B \to \infty} \left[ \frac{\partial \ln l}{\partial \mu} \right] = 0
\]  
(38)

The \( \ln l(\beta, \sigma^2, \lambda; y) \) for the N-HN, the MLEs=OLS (Waldman, 1982). First allow \( B \to \infty \) then set \( \mu = 0 \) and then allow \( \lambda \to 0 \) in the FOCs above. Note that the FOCs for N-HN model parameterizations do not have any indeterminate terms when \( \lambda \to 0 \), hence the FOC conditions are relatively easier to evaluate in the limit when compare to N-DTN and N-TN models above.

Recall: \( A_1 = \frac{(B-\mu)(\lambda^{-2}+1)^{\frac{3}{4}}}{\sigma}, \ A_2 = -\mu(\lambda^{-2}+1)^{\frac{1}{4}}, \ A_3 = \frac{(B+(y-x')\beta)\lambda+(B-\mu)\lambda^{-1}}{\sigma} \) and \( A_4 = \frac{(y-x'\beta)\lambda-\mu\lambda^{-1}}{\sigma} \).

For the N-HN model, \( A_4 = \frac{(y-x'\beta)\lambda}{\sigma} \) is the only relevant term after \( B \to \infty \) then set \( \mu = 0 \).

\[
\lim_{\lambda \to 0} \left[ \frac{\partial \ln l}{\partial \beta} \right]_{B \to \infty} = \hat{\beta}_{ols}
\]  
(39)
\[
\lim_{\lambda \to 0} \left|_{\mu = 0} \right. \lim_{B \to -\infty} \frac{\partial \ln l}{\partial \sigma^2} = \hat{\sigma}_v^2
\]

(40)

\[
\lim_{\lambda \to 0} \left|_{\mu = 0} \right. \lim_{B \to -\infty} \frac{\partial \ln l}{\partial \lambda} = 0
\]

(41)

C.0.2 SOC for N-DTN, N-TN and N-HN models

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\frac{\Sigma_i x_i x_i'}{\sigma^2} - \frac{\lambda^2}{\sigma^2} \Sigma_i \left[ \frac{\phi_i(A_3) A_3 - \phi_i(A_4) A_4}{[\Phi(A_3) - \Phi(A_4)]} \right] + \frac{(\phi_i(A_3) - \phi_i(A_4))^2}{[\Phi(A_3) - \Phi(A_4)]^2} x_i x_i'
\]

(42)

\[
\frac{\partial^2 \ln L}{\partial \mu \partial \beta} = \frac{\Sigma_i x_i}{\sigma^2} + \frac{\Sigma_i}{\sigma^2} \left[ \frac{-A^*_3 \phi_i(A_3) A_3 \frac{\partial A_3}{\partial \beta} + A^*_4 \phi_i(A_4) A_4 \frac{\partial A_4}{\partial \beta}}{[\Phi(A_3) - \Phi(A_4)]} \right] - \frac{(A^*_3 \phi_i(A_3) - A^*_4 \phi_i(A_4)) (\phi_i(A_3) \frac{\partial A_3}{\partial \beta} - \phi_i(A_4) \frac{\partial A_4}{\partial \beta})}{[\Phi(A_3) - \Phi(A_4)]^2} \text{ where } A^*_i = \frac{\partial A^*_i (i = 3,4)}{\partial \sigma}
\]

(43)

\[
\frac{\partial^2 \ln L}{\partial B \partial \beta} = \Sigma_i \left[ -A^*_3 \phi_i(A_3) A_3 \frac{\partial A_3}{\partial \beta} + \phi_i(A_3) \frac{\partial A^*_3}{\partial \beta} + \phi_i(A_4) \frac{\partial A^*_4}{\partial \beta} \right] - \frac{\phi_i(A_3) A_3 [\phi_i(A_3) \frac{\partial A_3}{\partial \beta} + \phi_i(A_4) \frac{\partial A_4}{\partial \beta}]}{[\Phi(A_3) - \Phi(A_4)]^2} \text{ where } A^*_i = \frac{\partial A^*_i (i = 3)}{\partial B}
\]

(44)

\[
\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta} = -\frac{\Sigma_i (y_i - x_i \beta + \mu) x_i}{\sigma^4} + \Sigma_i \left[ -A^*_3 \phi_i(A_3) A_3 \frac{\partial A_3}{\partial \beta} + \phi_i(A_3) \frac{\partial A^*_3}{\partial \beta} + A^*_4 \phi_i(A_4) A_4 \frac{\partial A_4}{\partial \beta} - \phi_i(A_4) \frac{\partial A^*_4}{\partial \beta} \right] - \frac{(\phi_i(A_3) A_3 - \phi_i(A_4) A_4) [\phi_i(A_3) \frac{\partial A_3}{\partial \beta} - \phi_i(A_4) \frac{\partial A_4}{\partial \beta}]}{[\Phi(A_3) - \Phi(A_4)]^2} \text{ where } A^*_i = \frac{\partial A^*_i (i = 3,4)}{\partial \sigma^2}
\]

(45)
\[ \frac{\partial^2 \ln L}{\partial \lambda \partial \beta} = \sum_i \left[ \phi(A_3)\left( \frac{\partial A_i^*}{\partial \lambda} - A_i^* \frac{\partial A_i}{\partial \lambda} A_3 \right) - \phi(A_4)\left( \frac{\partial A_i^*}{\partial \lambda} - A_i^* \frac{\partial A_i}{\partial \lambda} A_4 \right) \right] \frac{\frac{\partial^2}{\partial \lambda^2} - A_i^* \frac{\partial^2 A_i}{\partial \lambda^2} A_3}{\Phi(A_3) - \Phi(A_4)} \] where \( A_i^* = \frac{\partial A_i^*}{\partial \lambda} \) (46)

\[ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} = -n \left[ \frac{\partial A_3}{\partial \lambda} \phi(A_1) \frac{\partial A_i}{\partial \lambda} \frac{\partial A_i}{\partial B} A_1 \right] \frac{\Phi(A_1) - \Phi(A_2)}{\Phi(A_3) - \Phi(A_4)} + \sum_i \left[ \phi(A_3)\left( \frac{\partial A_i^*}{\partial \lambda} - A_i^* \frac{\partial A_i}{\partial \lambda} A_3 \right) \right] \frac{\frac{\partial^2}{\partial \lambda^2} - A_i^* \frac{\partial^2 A_i}{\partial \lambda^2} A_3}{\Phi(A_3) - \Phi(A_4)} \] where \( A_i^*, (i = 1, 2, 3 \text{ and } 4) = \frac{\partial A_i^*}{\partial \lambda} \) (47)

\[ \frac{\partial^2 \ln L}{\partial \mu \partial B} = -n \left[ \phi(A_1)\left( \frac{\partial A_i}{\partial \lambda} \frac{\partial A_i}{\partial \mu} A_1 \right) - \phi(A_1)\left( \frac{\partial A_i}{\partial \lambda} A_i^* \frac{\partial A_i^*}{\partial \mu} \right) \right] \frac{\Phi(A_1) - \Phi(A_2)}{\Phi(A_3) - \Phi(A_4)} + \sum_i \left[ \phi(A_3)\left( -A_i^* \frac{\partial A_i}{\partial \lambda} A_3 \right) \frac{\Phi(A_3) - \Phi(A_4)}{\Phi(A_3) - \Phi(A_4)} \right] \frac{\frac{\partial^2}{\partial \lambda^2} - A_i^* \frac{\partial^2 A_i}{\partial \lambda^2} A_3}{\Phi(A_3) - \Phi(A_4)} \] where \( A_i^*, (i = 1, 2, 3 \text{ and } 4) = \frac{\partial A_i^*}{\partial \lambda} \) (48)

\[ \frac{\partial^2 \ln L}{\partial \lambda \partial B} = -n \left[ \phi(A_1)\left( -A_i^* \frac{\partial A_i}{\partial \lambda} A_1 + \frac{\partial A_i}{\partial \lambda} A_1 \right) \frac{\Phi(A_1) - \Phi(A_2)}{\Phi(A_3) - \Phi(A_4)} \right] + \sum_i \left[ \phi(A_3)\left( -A_i^* \frac{\partial A_i}{\partial \lambda} A_3 + \frac{\partial A_i}{\partial \lambda} A_3 \right) \frac{\Phi(A_3) - \Phi(A_4)}{\Phi(A_3) - \Phi(A_4)} \right] \frac{\frac{\partial^2}{\partial \lambda^2} - A_i^* \frac{\partial^2 A_i}{\partial \lambda^2} A_3}{\Phi(A_3) - \Phi(A_4)} \] where \( A_i^*, (i = 1, 2, 3 \text{ and } 4) = \frac{\partial A_i^*}{\partial \lambda} \) (49)

\[ \frac{\partial^2 \ln L}{\partial B^2} = -n \left[ \left( -\phi(A_1) A_i^* A_1 \right) - \phi(A_1) A_i^* \right] \frac{\Phi(A_1) - \Phi(A_2)}{\Phi(A_3) - \Phi(A_4)} + \sum_i \left[ \phi(A_3)\left( -A_i^* A_i^* A_3 \right) \frac{\Phi(A_3) - \Phi(A_4)}{\Phi(A_3) - \Phi(A_4)} \right] \frac{\frac{\partial^2}{\partial \lambda^2} - A_i^* \frac{\partial^2 A_i}{\partial \lambda^2} A_3}{\Phi(A_3) - \Phi(A_4)} \] where \( A_i^* = \frac{\partial A_i}{\partial \lambda} \) and \( A_3^* = \frac{\partial A_3}{\partial \lambda} \) (50)
\[
\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^2} - n[\frac{\phi(A_1)(-A_1^2 A_1) + \phi(A_2)A_2^2 A_2}{\Phi(A_1) - \Phi(A_2)} - \frac{(\phi(A_1)A_1^* - \phi(A_2)A_2^*)^2}{[\Phi(A_1) - \Phi(A_2)]^2}]
+ \sum_i^{\eta} \left\{ \frac{\phi(A_3)(-A_3^2 A_3) + \phi(A_4)(A_4^2 A_4)}{[\Phi(A_3) - \Phi(A_4)]} - \frac{(\phi(A_3)A_3^* - \phi(A_4)A_4^*)^2}{[\Phi(A_3) - \Phi(A_4)]^2} \right\}, \quad A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \mu}
\]

(51)

\[
\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} = \frac{n}{\sigma^4} \left( y - x' \beta + \mu \right) - n[\frac{\phi(A_1)(\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1) - \phi(A_2)(\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2)}{[\Phi(A_1) - \Phi(A_2)]^2}]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
\]

where \( A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \sigma^2} \)

(52)

\[
\frac{\partial^2 \ln L}{\partial \lambda \partial \mu} = -n[\frac{\phi(A_1)(\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1) - \phi(A_2)(\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2)}{[\Phi(A_1) - \Phi(A_2)]^2}]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
\]

where \( A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \lambda} \)

(53)

\[
\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = \frac{n}{\sigma^4} \left( y - x' \beta + \mu \right) - n[\frac{\phi(A_1)(\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1) - \phi(A_2)(\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2)}{[\Phi(A_1) - \Phi(A_2)]^2}]
+ \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
- \sum_i^{\eta} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \right]
\]

where \( A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \sigma^2} \)

(54)
\[ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = -n \left[ \frac{\phi(A_1)(\frac{\partial A_1^*}{\partial \lambda} - A_1^* \frac{\partial A_1}{\partial \lambda}) - \phi(A_2)(\frac{\partial A_2^*}{\partial \lambda} - A_2^* \frac{\partial A_2}{\partial \lambda})}{[\Phi(A_1) - \Phi(A_2)]} \right] \\
+ \sum_{i=1}^{n} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \lambda} - A_3^* \frac{\partial A_3}{\partial \lambda}) - \phi(A_4)(\frac{\partial A_4^*}{\partial \lambda} - A_4^* \frac{\partial A_4}{\partial \lambda})}{[\Phi(A_3) - \Phi(A_4)]} \right] \quad \text{where } A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \sigma^2} \quad (55) \]

\[ \frac{\partial^2 \ln L}{\partial \lambda^2} = -n \left[ \frac{\phi(A_1)(\frac{\partial A_1^*}{\partial \lambda} - A_1^* \frac{\partial A_1}{\partial \lambda}) - \phi(A_2)(\frac{\partial A_2^*}{\partial \lambda} - A_2^* \frac{\partial A_2}{\partial \lambda})}{[\Phi(A_1) - \Phi(A_2)]^2} \right] \\
+ \sum_{i=1}^{n} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \lambda} - A_3^* \frac{\partial A_3}{\partial \lambda}) - \phi(A_4)(\frac{\partial A_4^*}{\partial \lambda} - A_4^* \frac{\partial A_4}{\partial \lambda})}{[\Phi(A_3) - \Phi(A_4)]^2} \right] \quad \text{where } A_i^*, (i = 1, \ldots, 4) = \frac{\partial A_i^*}{\partial \lambda} \quad (56) \]

Important notations that is used below for convenience.

Let:

\[ D = [\Phi(A_1) - \Phi(A_2)]^2 = \Phi^2(A_1) - 2\Phi(A_1)\Phi(A_2) + \Phi^2(A_2) \]

\[ \Rightarrow \frac{\partial D}{\partial \lambda} = 2\Phi(A_1)\phi(A_1)\frac{\partial A_1}{\partial \lambda} - 2[\Phi(A_2)\phi(A_1)\frac{\partial A_1}{\partial \lambda} + \Phi(A_1)\phi(A_2)\frac{\partial A_2}{\partial \lambda} + 2\Phi(A_2)\phi(A_2)\frac{\partial A_2}{\partial \lambda}] \]

\[ \Rightarrow \lim_{\lambda \to 0} \frac{\partial D}{\partial \lambda} = \phi(A_2)\frac{\partial A_2}{\partial \lambda}2[\Phi(A_1)\phi(A_1)\frac{\partial A_1}{\partial \lambda} - \Phi(A_2)\phi(A_1)\frac{\partial A_1}{\partial \lambda} - \Phi(A_1)\phi(A_2)\frac{\partial A_2}{\partial \lambda} - \Phi(A_2)\phi(A_2)\frac{\partial A_2}{\partial \lambda}] \]
$$\Rightarrow \frac{\partial D}{\partial \lambda} = -\Phi(A_1) + \Phi(A_2)$$

These terms will be the only relevant terms since $\phi(A_2) \frac{\partial A_2}{\partial \lambda}$ will cancel out with the expression in the numerator.

1. $\frac{\partial^2 D}{\partial \lambda^2} = -\phi(A_1) \frac{\partial A_1}{\partial \lambda} + \phi(A_2) \frac{\partial A_2}{\partial \lambda} = \phi(A_2) \frac{\partial A_2}{\partial \lambda} \left[ -\frac{\phi(A_1)}{\phi(A_2)} \frac{\partial A_1}{\partial \lambda} + 1 \right]$

Note when $D = [\Phi(A_3) - \Phi(A_4)]^2$

2. Use Fact 1 and Fact 2: $\frac{\phi(A_1)}{\phi(A_2)} \to 0$ and $\frac{\phi(A_3)}{\phi(A_4)} \to 0$ as $\lambda \to 0$ for $\mu \leq 0$

3. $\Sigma_{i=1}^n e_i = 0$ (the sum of the residuals equal to zero)

4. $\Sigma_{i=1}^n e_i x_i = 0$

To save on space I will use $\frac{\phi(A_1)}{\phi(A_2)}$ and $\frac{\phi(A_3)}{\phi(A_4)}$ for notational convenience.

C.0.3 This section shows all the proof of the SOC for N-DTN, N-TN and N-HN models when

$$\lambda \to 0$$

For N-DTN the SOCs when $\lambda \to 0$: 
\[
\lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \partial \beta} = -\frac{\Sigma_i x_i x_i'}{\sigma^2}; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \partial \sigma^2} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \sigma \lambda} = \frac{\Sigma_i x_i}{\sigma^2} \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \sigma \beta} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \sigma \mu} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \beta \mu \lambda} = 0; \\
\lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = -\frac{n}{2 \sigma}; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} = 0; \\
\lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \mu \lambda} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \lambda} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \mu} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \mu \mu} = 0; \\
\lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \sigma^2} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} = 0; \\
\lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \partial B} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \partial \mu} = 0; \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \mu \partial B} = 0.
\]

For N-TN the SOCs when \( \lambda \to 0 \):

\[
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \beta \partial \beta} \right) = -\frac{\Sigma_i x_i x_i'}{\sigma^2}; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \beta \partial \sigma^2} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \beta \sigma \lambda} \right) = \frac{\Sigma_i x_i}{\sigma^2} \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \beta \sigma \beta} \right) = 0; \\
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} \right) = -\frac{n}{2 \sigma}; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} \right) = 0; \\
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \mu \lambda} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \lambda \lambda} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \lambda \mu} \right) = 0; \\
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \mu \mu} \right) = 0; \\
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \lambda \sigma^2} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} \right) = 0; \\
\lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \lambda \partial B} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \lambda \partial \mu} \right) = 0; \lim_{\lambda \to 0} \lim_{B \to \infty} \left( \frac{\partial^2 \ln L}{\partial \mu \partial B} \right) = 0.
\]

For N-HN the SOCs when \( \lambda \to 0 \):

For the N-HN. First allow \( B \to \infty \) and then set \( \mu = 0 \) and then allow \( \lambda \to 0 \) in the SOCs above. Note that the SOCs for N-HN parameterization does not have any indeterminate terms when \( \lambda \to 0 \), hence the SOC conditions are relatively easier when compared to N-DTN and N-TN models above. For the (N-HN) only \( A_4 = \frac{(y-x')^2}{\sigma} \) is the relevant term after we substitute all the limits above.
\[
\lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial \beta \partial \beta} \right|_{\lambda} = -\frac{\sum_i x_i x_i'}{\sigma^2}; \lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial \beta \partial \sigma^2} \right|_{\lambda} = 0;
\]

\[
\lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial \beta \partial \lambda} \right|_{\sigma} = \frac{2}{\sqrt{2\pi} \sigma} \sum_{i=1}^n x_i;
\]

\[
\lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} \right|_{\lambda} = -\frac{n}{2\pi}; \lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} \right|_{\sigma} = 0;
\]

\[
\lim_{\lambda \to 0} \mu = 0 \lim_{\beta \to \infty} \left. \frac{\partial^2 \ln L}{\partial \lambda^2} \right|_{\sigma} = -\frac{2n}{\pi}
\]

C.0.4 Proof of all the SOCs above:

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \beta} = -\frac{\sum_i x_i x_i'}{\sigma^2} - \frac{\lambda^2 \sum_i [(\phi_i(A_3)A_3 - \phi_i(A_4)A_4) + (\phi_i(A_3) - \phi_i(A_4))^2]}{[\Phi(A_3) - \Phi(A_4)]^2} x_i x_i' \]

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \lambda} = \frac{\lambda^2 \sum_i [(\phi_i(A_3)A_3 - \phi_i(A_4)A_4) + (\phi_i(A_3) - \phi_i(A_4))^2]}{[\Phi(A_3) - \Phi(A_4)]^2} x_i x_i' \]

is indeterminate when \( \lambda \to 0 \)

Term1 \( = \frac{\lambda^2}{\sigma^2} \frac{\phi_i(A_3)A_3 + \phi_i(A_4)A_4}{[\Phi(A_3) - \Phi(A_4)]} = 0 \), when \( \lambda \to 0 \)

Term2 \( = \frac{\lambda^2}{\sigma^2} \frac{(\phi_i(A_3) - \phi_i(A_4))^2}{[\Phi(A_3) - \Phi(A_4)]^2} = 0 \), when \( \lambda \to 0 \)

let \( N = [A_3 A_3^2 \phi(A_3) - A_4 A_4^2 \phi(A_4)] \), let \( A_3^* = A_4^* = \frac{\lambda}{\sigma} \)

let \( D = [\Phi(A_3) - \Phi(A_4)] \)

For Term1 we can ignore the first term \((A_3 A_3^2 \phi(A_3))\) since it will be zero (using fact 2 above),

\[
\frac{\partial N}{\partial \lambda} = -\phi(A_4)[-A_3 A_4^2 \frac{\partial A_4}{\partial \lambda} + A_4 A_3 A_4^2 \frac{\partial A_4}{\partial \lambda} + A_4^2 \frac{\partial A_4}{\partial \lambda}]
\]

Let \( D = [\Phi(A_3) - \Phi(A_4)] \)

\[
\frac{\partial D}{\partial \lambda} = \phi(A_4) \frac{\partial A_4}{\partial \lambda} \frac{\phi(A_4) \frac{\partial A_4}{\partial \lambda} - 1}
\]
For Term 2:

let $N = A_4^2 \phi_i^2(A_4)$ since $\frac{\phi_i(A_4)}{\phi_i(A_4)} = 0$ in the limit so we can ignore $\phi_i(A_4)$.

\[
\frac{\partial N}{\partial \lambda} = \left[ \phi_i^2(A_4)2A_4 \frac{\partial A_4^2}{\partial \lambda} - 2A_4^2 A_4 \phi_i^2(A_4) \frac{\partial A_4}{\partial \lambda} \right] = \phi_i^2(A_4)2\left[ A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4^2 A_4 \frac{\partial A_4}{\partial \lambda} \right]
\]

Let $D = [\Phi(A_3) - \Phi(A_4)]^2$ after some algebra:

\[
\frac{\partial N}{\partial \lambda} = \frac{\phi_i(A_4)[A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4^2 A_4]}{[\Phi(A_3)\frac{\phi(A_3)}{\phi(A_4)}]_{\lambda=0} - (\Phi(A_4)\frac{\phi(A_3)}{\phi(A_4)} + \Phi(A_3)) + \Phi(A_4)]} = 0 \quad \text{as} \quad \lambda \rightarrow 0 \quad \text{do L'Hopital rule again}
\]

$\frac{\partial D}{\partial \lambda}$ is coming from 1(ii) above

let: $\hat{A}_4 = \frac{\partial A_4}{\partial \lambda}$

\[
\Rightarrow N = \phi_i(A_4)[A_4^2 \hat{A}_4 - A_4^2 A_4]
\]

\[
\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \phi_i(A_4)\left[ A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda} \right] - \frac{\partial A_4}{\partial \lambda} (A_4^2 \hat{A}_4 A_4 - A_4^2 A_4^2)
\]

\[
\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \frac{\phi_i(A_4)[A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda}]}{\phi(A_4)\frac{\phi(A_3)}{\phi(A_4)} + 1]}
\]

\[
\Rightarrow \text{Term 2} = [A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - 2A_4 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 - A_4^2 \hat{A}_4 A_4 + A_4^2 A_4^2]
\]
Term 1 = \(-A_2^2A_4^2 + 2A_4A_3^* \frac{\partial A_3^*}{\partial A_4} + A_4^2\)

Term 1 + Term 2 = \([A_4 \frac{\partial A_4}{\partial A_4} + \widehat{A}_4 \frac{\partial A_4}{\partial A_4} - A_4^* \widehat{A}_4 A_4]\)

Note that \(A_4^2 = \frac{1}{2} \sigma \frac{\partial A_4}{\partial A_4} = \frac{1}{\sigma} \frac{\partial A_4}{\partial A_4} = \frac{\varepsilon + \mu \lambda^{-2}}{\sigma} = \frac{\varepsilon \lambda^2 + \mu}{\sigma} \Rightarrow \widehat{A}_4 = \frac{\partial A_4}{\partial A_4} = \frac{1}{\varepsilon \lambda^2 + \mu} = \frac{\lambda^2}{\sigma(\varepsilon \lambda^2 + \mu)}\)

for \(A_4 \frac{\partial A_4}{\partial A_4} = (\frac{\varepsilon \lambda^2 + \mu}{\sigma}) \frac{\partial A_4}{\partial A_4} = (\frac{\varepsilon \lambda^2 + \mu}{\sigma}) \frac{\partial A_4}{\partial A_4} = (\frac{\varepsilon \lambda^2 + \mu}{\sigma}) \frac{\partial A_4}{\partial A_4} = 0 \text{ for } \lambda \rightarrow 0\)

Also note that

\[
\widehat{A}_4 \frac{\partial A_4}{\partial A_4} = \widehat{A}_4^2 = \left(\frac{\lambda^2}{\sigma(\varepsilon \lambda^2 + \mu)}\right)^2 = 0 \text{ for } \lambda \rightarrow 0
\]

\(A_4^* \widehat{A}_4 A_4 = \left(\frac{\lambda}{\sigma}\right) \left(\frac{\lambda^2}{\sigma(\varepsilon \lambda^2 + \mu)}\right) \left(\frac{\varepsilon \lambda^2 + \mu}{\sigma}\right) = 0 \text{ for } \lambda \rightarrow 0\)

\[
\Rightarrow \text{ Term 1 + Term 2} = 0 \text{ for } \lambda \rightarrow 0
\]

\[
\Rightarrow \lim_{\lambda \rightarrow 0} \left[\frac{\partial^2 \ln L}{\partial \beta \partial \beta'}\right] = -\frac{\Sigma_i x_i x'_i}{\sigma^2} (57)
\]

End Proof for SOC for \((\frac{\partial^2 \ln L}{\partial \beta \partial \beta'})\).
Proof:

when \( \lambda \to 0 \) \( \Rightarrow \left\{ \frac{-A_3^4 \phi_i(A_3) A_3 \frac{\partial A_i}{\partial a} + A_3^4 \phi_i(A_3) A_4 \frac{\partial A_i}{\partial a}}{[\Phi(A_3) - \Phi(A_4)]} \right\} = 0 \) \( \Rightarrow \frac{\Phi(A_3) A_3 \frac{\partial A_i}{\partial a} + A_3^2 \phi_i(A_3) [\Phi(A_3) - \Phi(A_4)]}{[\Phi(A_3) - \Phi(A_4)]^2} \) = \( \frac{0}{0} \). 

Note that \( \phi_i(A_3) \) is not important in the limit so it can be ignored.

\[
\text{Term1} = \frac{A_4^4 \phi_i(A_4) A_4 \frac{\partial A_i}{\partial a}}{[\Phi(A_3) - \Phi(A_4)]}
\]

note that \( A_4 A_4^* \frac{\partial a}{\partial a} = A_4 \hat{A}_4 \), where \( \hat{A}_4 = \frac{\partial a^4}{\partial a} A_4^* \).

let \( N = [\phi_i(A_4) A_4 \hat{A}_4] \Rightarrow \frac{\partial N}{\partial a} = \phi_i(A_4) [\hat{A}_4 \frac{\partial A_i}{\partial a} + A_4 \frac{\partial \hat{A}_4}{\partial a} - A_4^2 \hat{A}_4 \frac{\partial A_i}{\partial a}] \)

\( \Rightarrow \frac{\partial N}{\partial a} = \phi_i(A_4) [\hat{A}_4 \frac{\partial A_i}{\partial a} + A_4 \frac{\partial \hat{A}_4}{\partial a} - A_4^2 \hat{A}_4 \frac{\partial A_i}{\partial a}] = \phi_i(A_4) [\hat{A}_4 \frac{\partial A_i}{\partial a} - A_4^2 \hat{A}_4 \frac{\partial A_i}{\partial a}] \) because \( \frac{\partial \hat{A}_4}{\partial a} = 0 \)

\( \Rightarrow \frac{\partial N}{\partial a} = \frac{[\hat{A}_4 - A_4^2 \hat{A}_4]}{\phi_i(A_4) \frac{\partial a}{\partial a} - 1} = [-\hat{A}_4 + A_4^2 \hat{A}_4] \) in the limit.

After some algebra:

\[
\text{Term1} = [-\hat{A}_4 + A_4^2 \hat{A}_4]
\]

Let

\[
\text{Term2} = \frac{A_4 \frac{\partial A_i}{\partial a} \phi_i^2(A_4)}{[\Phi(A_3) - \Phi(A_4)]^2}
\]

\[
\text{Term2} = \frac{A_4 \frac{\partial A_i}{\partial a} \phi_i^2(A_4)}{[\Phi(A_3) - \Phi(A_4)]^2} \]

\( \Rightarrow \frac{\partial N}{\partial a} = \phi_i(A_4) [-2 A_4 \hat{A}_4 \frac{\partial A_i}{\partial a} + \frac{\partial \hat{A}_4}{\partial a}] = \phi_i(A_4) [-2 A_4 \hat{A}_4 \frac{\partial A_i}{\partial a}] \)

and \( \frac{\partial D}{\partial a} = \phi(A_4) \frac{\partial A_i}{\partial a} \phi(A_3) \frac{\partial A_i}{\partial a} \frac{\partial A_i}{\partial a} - \left[ \Phi(A_4) \frac{\phi(A_3) \partial A_i}{\phi(A_4)} \right] + \Phi(A_4) \] using 1(ii) above.

\[
\Rightarrow \frac{\partial N}{\partial a} = \phi(A_4) \frac{[-A_4 \hat{A}_4]}{[\Phi(A_3) \phi(A_3) \frac{\partial A_i}{\partial a} - \left[ \Phi(A_4) \frac{\phi(A_3) \partial A_i}{\phi(A_4)} \right] + \Phi(A_4)]} = \frac{0}{0}
\]
\[
\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = -\phi_i(A_4)\left(\hat{A}_1 \frac{\partial A_4}{\partial \lambda} + A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda}\right)
\]

\[
\Rightarrow \frac{\partial^2 N}{\partial \theta^2} = \phi_i(A_4)\left[\hat{A}_4 \frac{\partial A_4}{\partial \lambda} + A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda}\right] - \frac{\phi_i(A_4)\frac{\partial A_4}{\partial \lambda} + 1}{\phi_i(A_4)\frac{\partial \hat{A}_4}{\partial \lambda}} = -[\hat{A}_4 - A_4^2 \hat{A}_4]
\]

After some Algebra:

\[\text{Term 2} = -[\hat{A}_4 - A_4^2 \hat{A}_4]\]

\[\Rightarrow \text{Term 1 - Term 2} = \left[-\hat{A}_4 + A_4^2 \hat{A}_4\right] - \left[-\hat{A}_4 - A_4^2 \hat{A}_4\right] = 0\]

\[\Rightarrow \lim_{\lambda \to 0} \left|\frac{\partial^2 \ln l}{\partial \mu \partial \beta}\right| = \Sigma_i x_i \sigma^2\]

(58)

End Proof of SOC for \(\frac{\partial^2 \ln l}{\partial \mu \partial \beta}\).

\[
\frac{\partial^2 \ln L}{\partial B \partial \beta} = \sum_i \left(\frac{-A_3^2 \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \theta} + \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \lambda}}{[\Phi(A_3) - \Phi(A_4)]} - \frac{\phi_i(A_3)A_3 \frac{\partial A_4}{\partial \theta} + \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \lambda}}{[\Phi(A_3) - \Phi(A_4)]^2}\right), A_i^* = \frac{\partial A_i}{\partial B}(i = 3)
\]

Term 1 = \(-A_3^2 \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \theta} + \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \lambda}\) = 0, as \(\lambda \to 0\)

Term 2 = \(\frac{\phi_i(A_3)A_3 \frac{\partial A_4}{\partial \theta} + \phi_i(A_3)A_3 \frac{\partial A_4}{\partial \lambda}}{[\Phi(A_3) - \Phi(A_4)]^2}\) = 0, as \(\lambda \to 0\)

Proof: Note that \(A_3^* \frac{\partial A_4}{\partial \beta} = \frac{\lambda + \lambda^{-1}}{\sigma} \frac{-x_3}{\sigma} = \frac{\lambda^2 + 1}{\sigma} = \hat{A}_4\). For Term 2 let \(N = \phi_i(A_3)\phi_i(A_4)\hat{A}_4\)

\[\Rightarrow \frac{\partial N}{\partial \lambda} = \hat{A}_4 \phi_i(A_4) \phi_i(A_3) \frac{\partial A_3}{\partial \lambda} + \phi_i(A_3) \phi_i(A_4) \frac{\partial A_4}{\partial \lambda} + \phi_i(A_3) \phi_i(A_4) \frac{\partial \hat{A}_4}{\partial \lambda}\]

and
\[ \frac{\partial N}{\partial \lambda} = \phi(\lambda) \frac{\partial A_4}{\partial \lambda} 2[\Phi(A_3) \phi(A_3) \frac{\partial A_3}{\partial \lambda} - [\Phi(A_4) \phi(A_3) \frac{\partial A_3}{\partial \lambda} + \Phi(A_3)] + \Phi(A_4)] \]

After some algebra

\[ \frac{\partial N}{\partial \lambda} = \frac{\widehat{A}_4[\phi'(A_3) \frac{\partial A_3}{\partial \lambda} - \phi_1(A_3) A_4]}{2[\Phi(A_3) \phi(A_3) \frac{\partial A_3}{\partial \lambda} - [\Phi(A_4) \phi(A_3) \frac{\partial A_3}{\partial \lambda} + \Phi(A_3)] + \Phi(A_4)]} = 0 \text{ as } \lambda \to 0 \]

Apply L'Hospital rule again:

\[ \Rightarrow \text{Term 2} = \frac{\partial^2 N}{\partial \sigma^2 \partial \beta} = \frac{\widehat{A}_4[\phi'(A_3) \frac{\partial A_3}{\partial \lambda} - \phi_1(A_3) A_4]}{2[\Phi(A_3) \phi(A_3) \frac{\partial A_3}{\partial \lambda} - [\Phi(A_4) \phi(A_3) \frac{\partial A_3}{\partial \lambda} + \Phi(A_3)] + \Phi(A_4)]} = 0 \text{ since } \phi_1(A_3) = 0, \lambda \to 0 \]

\[ \Rightarrow \lim_{\lambda \to 0} \frac{\partial^2 \ln l}{\partial B \partial \beta} = 0 \quad (59) \]

End Proof of SOC for \[ \frac{\partial^2 \ln L}{\partial B \partial \beta} \].

\[ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta} = - \frac{\sum_i [y_i - x_i \beta + \mu x_i]}{\sigma^4} + \sum_i \left[ -A_i^2 \phi(A_3) A_3 \frac{\partial A_3}{\partial \sigma} + \phi_1(A_3) \frac{\partial A_3}{\partial \beta} + A_i^4 \phi_1(A_4) A_4 \frac{\partial A_4}{\partial \sigma} - \phi_1(A_4) \frac{\partial A_4}{\partial \beta} \right] \]

\[ - \frac{[\phi_1(A_3) A_i^2 - \phi_1(A_4) A_i^1] [\phi_3(A_3) \frac{\partial A_3}{\partial \sigma} - \phi_1(A_4) \frac{\partial A_4}{\partial \beta}]^2}{[\Phi(A_3) - \Phi(A_4)]^2} \]

where \( A_i^* = \frac{\partial A_i^*}{\partial \sigma} \)

\[ \text{Term 1} = \frac{A_i^* \phi_1(A_4) A_4 \frac{\partial A_4}{\partial \sigma} - \phi_1(A_4) \frac{\partial A_4}{\partial \beta}}{[\Phi(A_3) - \Phi(A_4)]} = 0, \lambda \to 0 \]

\[ \text{Proof:} \quad \text{Note that} \quad A_4 A_3^2 \frac{\partial A_3}{\partial \beta} = A_4 \widehat{A}_4 \Rightarrow N = \phi_1(A_4) [A_4 \widehat{A}_4 - \frac{\partial A_4^*}{\partial \beta}] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi_1(A_4) [A_4 \frac{\partial \widehat{A}_4}{\partial \lambda} + \widehat{A}_4 \frac{\partial A_4}{\partial \lambda} - \frac{\partial A_4^*}{\partial \beta \partial \lambda} - A_i^2 \widehat{A}_4 \frac{\partial A_4}{\partial \lambda} + A_i^4 \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \lambda}] \]

81
\[
\frac{\partial N}{\partial \lambda} = \left[A_4 \frac{\partial A_4}{\partial \lambda} + \hat{A}_4 + A_4 \frac{\partial A_4^*}{\partial \beta} - A_4^2 \hat{A}_4 + A_4 \frac{\partial A_4^*}{\partial \beta} \right] \frac{[\phi(A_3) \frac{\partial A_3}{\partial \lambda} - 1]}{\phi(A_3) \frac{\partial A_3}{\partial \lambda} - 1}
\]

After some algebra Term 1:

\[
\text{Term 1} = -A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - \hat{A}_4 + A_4 \frac{\partial A_4^*}{\partial \lambda} + A_4^2 \hat{A}_4 - A_4 \frac{\partial A_4^*}{\partial \beta}
\]

Term 2 = \(-\phi_1^2(A_4)A_4 \frac{\partial A_4}{\partial \lambda}\) \Rightarrow \(N = [\phi_1^2(A_4) \hat{A}_4] = \frac{\partial N}{\partial \lambda} = \phi_1^2(A_4) - 2A_4 \frac{\partial A_4}{\partial \lambda} + A_4 \frac{\partial A_4^*}{\partial \beta}
\]

\[
\Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi_1^2(A_4) \frac{\partial A_4}{\partial \lambda} - 2A_4 \frac{\partial A_4}{\partial \lambda}}{\phi_1^2(A_4) \frac{\partial A_4}{\partial \lambda} - 2A_4 \frac{\partial A_4}{\partial \lambda} + A_4 \frac{\partial A_4^*}{\partial \beta}} = \phi_1(A_4) [\hat{A}_5 - A_4 \hat{A}_4],
\]

where \(\hat{A}_5 = 0.5 \frac{\partial A_4}{\partial \lambda}\)

\[
\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \phi_1(A_4) \left[ \frac{\partial \hat{A}_5}{\partial \lambda} - \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - \hat{A}_4 \frac{\partial A_4^*}{\partial \lambda} + A_4 \frac{\partial A_4^*}{\partial \beta} \right]
\]

\[
\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \left[ - \frac{\partial \hat{A}_5}{\partial \lambda} + \frac{\partial A_4}{\partial \lambda} \hat{A}_4 + \hat{A}_5 A_4 - A_4^2 \hat{A}_4 \right]
\]

After some algebra Term 2:

\[
\text{Term 2} = \left[ - \frac{\partial \hat{A}_5}{\partial \lambda} + \frac{\partial A_4}{\partial \lambda} \hat{A}_4 + \hat{A}_5 A_4 - A_4^2 \hat{A}_4 \right]
\]

\[
\Rightarrow \text{Term 1 + Term 2} = \left[ - \frac{\partial \hat{A}_5}{\partial \lambda} + \frac{\partial A_4^*}{\partial \lambda} + \hat{A}_5 A_4 - A_4 \frac{\partial A_4^*}{\partial \beta} \right] = \frac{x \mu}{2\sigma^2}
\]

\[
\Rightarrow \lim_{\lambda \to 0} \left[ \frac{\partial^2 \ln l}{\partial \sigma^2 \partial \beta} \right] = -\frac{\Sigma_i[n_i - x_i \beta + \mu] x_i}{\sigma^4} + \frac{\mu \Sigma_i^n x}{2\sigma^2} = 0 \text{ since } \Sigma_i^n c_i x_i = 0 \quad (60)
\]

End Proof of SOC for \(\frac{\partial^2 \ln l}{\partial \sigma^2 \partial \beta} \).
\[
\frac{\partial^2 \ln l}{\partial \lambda \partial \beta} = \sum_{i} \left[ \frac{\phi(A_3)(\frac{\partial A_3^*}{\partial \beta} - A_3^* \frac{\partial A_3}{\partial \beta}) - \phi(A_4)(\frac{\partial A_4^*}{\partial \beta} - A_4^* \frac{\partial A_4}{\partial \beta})}{[\Phi(A_3) - \Phi(A_4)]} \right] \frac{A_i}{[\Phi(A_3) - \Phi(A_4)]^2} \]

where \( A_i = \frac{\partial A_i^*(i=3,4)}{\partial \lambda} \)

\[
\text{Term1} = -\phi(A_4)(\frac{\partial A_4^*}{\partial \beta} - A_4^* \frac{\partial A_4}{\partial \beta}) = 0 \text{ for } \lambda \rightarrow 0
\]

\[
\text{Term 2} = -\phi_i^2(A_4) \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \beta} = 0 \text{, for } \lambda \rightarrow 0
\]

**Proof:** For Term1, note that \(-A_4 A_4^* \frac{\partial A_4}{\partial \beta} = -A_4 \hat{A}_4 \Rightarrow N = \phi_i(A_4)[-A_4 \hat{A}_4 + \frac{\partial A_4^*}{\partial \beta}]\)

\[
\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_4)[-A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - \hat{A}_4 \frac{\partial A_4}{\partial \lambda} + \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \beta} + A_4^2 \frac{\partial A_4}{\partial \beta} - A_4 \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \beta}]
\]

\[
\Rightarrow \frac{\partial N}{\partial \beta} = \left[-A_4 \frac{\partial \hat{A}_4}{\partial \beta} + \frac{\partial A_4^*}{\partial \beta} + A_4^2 \frac{\partial A_4}{\partial \beta} - A_4 \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \beta} - \frac{\phi_i(A_4)}{\Phi(A_4)} \frac{\partial \phi_i(A_4)}{\partial \beta} \frac{\partial A_4}{\partial \beta} - 1\right]
\]

\[
\Rightarrow \text{Term1} = [-A_4 \frac{\partial \hat{A}_4}{\partial \beta} - \hat{A}_4 + \frac{\partial A_4^*}{\partial \beta} + A_4^2 \frac{\partial A_4}{\partial \beta} - A_4 \frac{\partial A_4^*}{\partial \beta} \frac{\partial A_4}{\partial \beta}]
\]

For Term2:

\[
\Rightarrow N = [\phi_i^2(A_4) \hat{A}_4] = \frac{\partial N}{\partial \lambda} = \phi_i^2(A_4)[\frac{\partial \hat{A}_4}{\partial \lambda} - 2 \hat{A}_4 \frac{\partial A_4}{\partial \lambda}]
\]

\[
\Rightarrow \frac{\partial N}{\partial \beta} = \frac{\phi_i^2(A_4)[\frac{\partial \hat{A}_4}{\partial \lambda} - 2 \hat{A}_4 \frac{\partial A_4}{\partial \lambda}]}{2[A_4 \frac{\partial A_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - \Phi(A_3) \frac{\phi(A_3)}{\phi(A_4)} \frac{\partial A_4}{\partial \lambda} + \Phi(A_3)]} = \phi_i(A_4)[\hat{A}_5 - A_4 \hat{A}_4]
\]

where \( \hat{A}_5 = \frac{0.5 \frac{\partial A_4}{\partial \lambda}}{\lambda} \)
Proof: Note that

\[
\frac{\partial^2 N}{\partial \lambda^2} = \phi_1(A_4)[\frac{\partial A_5}{\partial \lambda} - \frac{\partial \hat{A}_4}{\partial \lambda} A_4 - \hat{A}_4 \frac{\partial A_5}{\partial \lambda} - \hat{A}_5 \frac{\partial A_4}{\partial \lambda} A_4 + A_4^2 \frac{\partial A_4}{\partial \lambda}] \\
\Rightarrow \frac{\partial^2 N}{\partial \lambda \partial \beta} = [-\frac{\partial A_5}{\partial \lambda} + \frac{\partial \hat{A}_4}{\partial \lambda} A_4 + \hat{A}_4 + \hat{A}_5 A_4 - A_4^2 \hat{A}_4]
\]

\[A_4^* = \frac{\varepsilon \mu \lambda^{-2}}{\sigma} \Rightarrow \frac{\partial A_4^*}{\partial \beta} = \frac{-\varepsilon}{\sigma} \Rightarrow \frac{\partial A_4^*}{\partial \lambda} = \frac{\varepsilon \lambda^2 + \mu}{\lambda^2 \sigma} \Rightarrow \frac{\partial A_4^*}{\partial \beta} = \frac{\partial A_4^*}{\partial \lambda} = 0 \text{ and } \frac{\partial A_4}{\partial \lambda} = \frac{\partial A_4^*}{\partial \beta} = 0
\]

\[\frac{\partial \hat{A}_4}{\partial \beta} = \frac{[(\varepsilon \lambda^2 + \mu)(2\varepsilon \lambda) - (\varepsilon \lambda^2 + \mu)(\varepsilon \lambda^2)]}{2\varepsilon (\varepsilon \lambda^2 + \mu)^2} = 0 \text{ and } \hat{A}_5 A_4 = -\frac{\varepsilon \lambda^2 - 2\varepsilon \lambda \mu + \frac{\mu^2}{\lambda^2}}{2\varepsilon \lambda^2 + \mu} = 0
\]

\[A_4 \frac{\partial A_4^*}{\partial \beta} = \frac{(\varepsilon \lambda + \frac{\mu}{\lambda})}{\sigma^2} = 0
\]

Term 2 = \[-\frac{\partial A_5}{\partial \lambda} + \frac{\partial \hat{A}_4}{\partial \lambda} A_4 + \hat{A}_4 + \hat{A}_5 A_4 - A_4^2 \hat{A}_4]\]

Term 1 + Term 2 = \[-\frac{\partial A_5}{\partial \lambda} + \frac{\partial \hat{A}_4}{\partial \lambda} A_4 + \hat{A}_4 + A_4 \frac{\partial A_4^*}{\partial \beta} = 0
\]

\[\Rightarrow \lim_{\lambda \to 0} \left| \frac{\partial^2 \ln l}{\partial \lambda \partial \beta} \right| = 0 \quad (61)
\]

End Proof of SOC for \[\frac{\partial^2 \ln L}{\partial \sigma \partial B}\].

\[\frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} = -n[\phi(A_1)(\frac{\partial A_1}{\partial B} - A_1 \frac{\partial A_1}{\partial B}) \frac{\phi(A_1)}{\Phi(A_1) - \Phi(A_2)}] - (\phi(A_1) A_1^* - \phi(A_2) A_2^*) \frac{\phi(A_1)}{\Phi(A_1) - \Phi(A_2)}^2 + \Sigma_i \left[\phi(A_i)(\frac{\partial A_i}{\partial B} - A_3 \frac{\partial A_i}{\partial B}) \frac{\phi(A_i)}{\Phi(A_3) - \Phi(A_4)} - \phi(A_3) A_3^* - \phi(A_i) A_i^*) \frac{\phi(A_i)}{\Phi(A_3) - \Phi(A_4)}^2 \right]
\]

where \[A_i^*, (i = 1, 2, 3 \text{ and } 4) = \frac{\partial A_i}{\partial \beta^2}\]

Proof: Note that \[\frac{\phi(A_1)}{\phi(A_2)} \to 0 \text{ and } \frac{\phi(A_1)}{\phi(A_3)} \to 0 \text{ in the limit then}
\]

\[\Rightarrow \lim_{\lambda \to 0} \left| \frac{\partial^2 \ln l}{\partial \sigma^2 \partial B} \right| = 0 \text{ for } \lambda \to 0 \quad (62)
\]
End Proof of SOC for $[\frac{\partial^2 \ln L}{\partial \mu \partial B}]$.

$$\frac{\partial^2 \ln L}{\partial \mu \partial B} = -n \left[ \frac{\phi(A_1) - A_1^* \frac{\partial A_1}{\partial B} A_1}{\Phi(A_1) - \Phi(A_2)} \right] - \left[ \frac{\phi(A_1) A_1^* - \phi(A_2) A_2^* \phi(A_1) \frac{\partial A_1}{\partial B}}{\Phi(A_1) - \Phi(A_2)} \right] + \sum_i \left[ \frac{\phi(A_3) - A_3^* \frac{\partial A_3}{\partial B} A_3}{\Phi(A_3) - \Phi(A_4)} \right] - \left[ \frac{\phi(A_3) A_3^* - \phi(A_4) A_4^* \phi(A_3) \frac{\partial A_3}{\partial B}}{\Phi(A_3) - \Phi(A_4)} \right]$$

**Proof:** Note that $\frac{\phi(A_1)}{\phi(A_2)} \to 0$ and $\frac{\phi(A_3)}{\phi(A_4)} \to 0$ in the limit then

$$\Rightarrow \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \mu \partial B} = 0, \lambda \to 0 \quad (63)$$

End Proof of SOC for $[\frac{\partial^2 \ln L}{\partial \mu \partial B}]$.

$$\frac{\partial^2 \ln L}{\partial \lambda \partial B} = -n \left[ \frac{\phi(A_1) - A_1^* \frac{\partial A_1}{\partial B} A_1}{\Phi(A_1) - \Phi(A_2)} \right] - \left[ \frac{\phi(A_1) A_1^* - \phi(A_2) A_2^* \phi(A_1) \frac{\partial A_1}{\partial B}}{\Phi(A_1) - \Phi(A_2)} \right] + \sum_i \left[ \frac{\phi(A_3) - A_3^* \frac{\partial A_3}{\partial B} A_3}{\Phi(A_3) - \Phi(A_4)} \right] - \left[ \frac{\phi(A_3) A_3^* - \phi(A_4) A_4^* \phi(A_3) \frac{\partial A_3}{\partial B}}{\Phi(A_3) - \Phi(A_4)} \right]$$

where $A_i^*, (i = 1, 2, 3$ and $4) = \frac{\partial A_i}{\partial \lambda}$

**Proof:** Note that $\frac{\phi(A_1)}{\phi(A_2)} \to 0$ and $\frac{\phi(A_3)}{\phi(A_4)} \to 0$ in the limit then

$$\Rightarrow \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \partial B} = 0, \lambda \to 0 \quad (64)$$

End Proof of SOC for $[\frac{\partial^2 \ln L}{\partial \lambda^2}]$.

$$\frac{\partial^2 \ln L}{\partial B^2} = -n \left[ \frac{\phi(A_1) - A_1^* \frac{\partial A_1}{\partial B} A_1}{\Phi(A_1) - \Phi(A_2)} \right] - \left[ \frac{\phi(A_1) A_1^* - \phi(A_2) A_2^* \phi(A_1) \frac{\partial A_1}{\partial B}}{\Phi(A_1) - \Phi(A_2)} \right] + \sum_i \left[ \frac{\phi(A_3) - A_3^* \frac{\partial A_3}{\partial B} A_3}{\Phi(A_3) - \Phi(A_4)} \right] - \left[ \frac{\phi^2(A_3) A_3^* \frac{\partial A_3}{\partial B}}{\Phi(A_3) - \Phi(A_4)} \right]$$

where $A_i^*, (i = 1, 3) = \frac{\partial A_i}{\partial B}$
Proof: Note that $\frac{\phi(A_1)}{\phi(A_2)} \to 0$ and $\frac{\phi(A_3)}{\phi(A_4)} \to 0$

$\Rightarrow \left[ \frac{\partial^2 \ln L}{\partial B^2} \right] = 0, \lambda \to 0 \quad (65)$

End Proof of SOC for $[\frac{\partial^2 \ln L}{\partial B^2}]$.

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^2} - n [\frac{\phi(A_1)(-A_1^2 A_1) + \phi(A_2)A_2^2 A_2}{[\Phi(A_1) - \Phi(A_2)]} - \frac{(\phi(A_1)A_1^* - \phi(A_2)A_2^*)^2}{[\Phi(A_1) - \Phi(A_2)]^2}]$$

$$+ \sum_i^n \frac{[\phi(A_3)(-A_3^2 A_3) + \phi(A_4)(A_4^* A_4)]}{[\Phi(A_3) - \Phi(A_4)]} - \frac{(\phi(A_1)A_1^* - \phi(A_2)A_2^*)^2}{[\Phi(A_3) - \Phi(A_4)]^2} \text{ where } A_i^*, (i=1,2,3,4)=\frac{\partial A_i^*}{\partial \mu}$$

Let:

- Term1 = $\frac{\phi(A_2)A_2^2 A_2}{[\Phi(A_1) - \Phi(A_2)]}$
- Term2 = $-\frac{\phi^2(A_2)A_2^2}{[\Phi(A_1) - \Phi(A_2)]^2}$
- Term3 = $\frac{\phi(A_4)(-A_4^2 A_4)}{[\Phi(A_3) - \Phi(A_4)]}$
- Term4 = $-\frac{\phi^2(A_4)A_4^*}{[\Phi(A_3) - \Phi(A_4)]^2}$

Proof:

For Term 1, let $N = \phi_i(A_2)[A_2^2 A_2]$

$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i(A_2)[A_2^2 \frac{\partial A_2}{\partial \lambda} + 2A_2 A_2^* \frac{\partial A_2^*}{\partial \lambda} - A_2^2 A_2^* \frac{\partial A_2}{\partial \lambda}]$

$$\frac{\partial N}{\partial \lambda} = \frac{[A_2^2 + 2A_2 A_2^* \frac{\partial A_2^*}{\partial \lambda} - A_2^2 A_2^*]}{[\Phi(A_2)]^2 \frac{\partial A_2^*}{\partial \lambda} - 1}$$

86
\[ \Rightarrow \text{Term 1} = [-A_2^2 - 2A_2 A_2' \frac{\partial A_2^*}{\partial \lambda} + A_2^* A_2'^*] \]

For Term 2= \( \frac{-\phi^2(A_2)A_2^*}{\varPsi(A_2)-\varPsi'(A_2)} \), let \( N = \phi^2_i(A_2)A_2^* \)

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi^2_i A_2 \left[ A_2' \frac{\partial A_i^*}{\partial \lambda} - A_2^* A_2' \frac{\partial A_2}{\partial \lambda} \right] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi^2_i A_2 \left[ \frac{\partial A_2}{\partial \lambda} \frac{\partial A_i^*}{\partial \lambda} - A_2^* \frac{\partial A_2}{\partial \lambda} - \frac{\partial A_i^*}{\partial \lambda} \right]}{\varPsi(A_2)-\varPsi'(A_2)} = \phi_i A_2 \left[ \frac{\partial A_i^*}{\partial \lambda} - A_2^* A_2' \right] = \phi_i A_2 \left[ A_2' \frac{\partial A_i^*}{\partial \lambda} - A_2^* A_2' \right] - A_2^* A_2' \]

\[ \Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \phi_i (A_2) \left[ A_2' \frac{\partial A_i^*}{\partial \lambda} + \frac{\partial A_2^*}{\partial \lambda} A_2' - \frac{\partial A_2}{\partial \lambda} A_2' - 2A_2 A_2' \frac{\partial A_2'}{\partial \lambda} + \frac{\partial A_i^*}{\partial \lambda} A_2 + A_2^* A_2' \frac{\partial A_2}{\partial \lambda} + A_2^* A_2' \frac{\partial A_2'}{\partial \lambda} \right] \]

\[ \Rightarrow \text{Term 2} = \frac{\partial^2 N}{\partial \lambda^2} \left[ -A_2' \frac{\partial A_i^*}{\partial \lambda} + \frac{\partial A_2^*}{\partial \lambda} A_2' + A_2^* + 2A_2 A_2' \frac{\partial A_2'}{\partial \lambda} + A_2^* A_2' - A_2^* A_2' \right] \]

Note that \( \frac{\partial A_i^*}{\partial \lambda} = \frac{1}{\mu} \Rightarrow \frac{\partial A_i^*}{\partial \lambda} = 0 \), \( -\hat{A}_2 \frac{\partial A_2^*}{\partial \lambda} = \frac{1}{\mu^2} \), and \( A_2^* \hat{A}_2 A_2 = \frac{1+\lambda^2}{\sigma^2} \).

After some algebra:

\[ \text{Term 1 + Term 2} = \left[ -\hat{A}_2 \frac{\partial A_i^*}{\partial \lambda} + A_2^* \hat{A}_2 A_2 \right] = \left[ \frac{1}{\mu^2} + \frac{1 + \lambda^2}{\sigma^2} \right] \]

Term 3= \( \frac{\phi(A_4)[-A_2^* A_4]}{[\varPsi(A_4)-\varPsi'(A_4)]} \Rightarrow \phi(A_4)A_i^* A_4 \)

\[ \Rightarrow N = \phi_i (A_4) [A_i^* A_4] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i (A_4) \left[ A_i^* \frac{\partial A_4}{\partial \lambda} + 2A_4 A_i^* \frac{\partial A_i^*}{\partial \lambda} - A_i^* A_4 \frac{\partial A_4}{\partial \lambda} \right] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi_i (A_4) \left[ A_i^* \frac{\partial A_4}{\partial \lambda} + 2A_4 A_i^* \frac{\partial A_i^*}{\partial \lambda} - A_i^* A_4 \frac{\partial A_4}{\partial \lambda} \right]}{\varPsi(A_4)-\varPsi'(A_4)} = \left[ A_i^* + 2A_4 A_i^* \frac{\partial A_i^*}{\partial \lambda} - A_i^* A_4 \frac{\partial A_4}{\partial \lambda} \right] = \left[ -A_i^* - 2A_4 A_i^* \frac{\partial A_i^*}{\partial \lambda} + A_i^* A_4 \right] \]
Term4 = \( -\phi^2(A_4)A_4^2 \)

Note that \( \phi(A_4)A_4^2 \Rightarrow N = \phi^2_1(A_4)A_4^2 \)

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi^2_1 A_4 2[A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4 A_4^3 \frac{\partial A_4}{\partial \lambda}] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi^2_1 A_4 2[A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4 A_4^3 \frac{\partial A_4}{\partial \lambda}]}{\phi(A_4) \frac{\partial A_4}{\partial \lambda} [\phi(A_4) \frac{\partial A_4}{\partial \lambda} + \Phi(A_4)]} \]

\[ \Rightarrow N = \phi_i A_4 [A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4 A_4^3 \frac{\partial A_4}{\partial \lambda}] = \phi_i(A_4)[A_4^2 A_4 - A_4^2 A_4] \]

\[ \Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \phi_i(A_4)[A_4^2 \frac{\partial A_4}{\partial \lambda} + A_4^3 \frac{\partial A_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda} - 2 A_4 A_4^3 \frac{\partial A_4}{\partial \lambda} + A_4^2 A_4^4 \frac{\partial A_4}{\partial \lambda} + A_4^2 A_4^2 \frac{\partial A_4}{\partial \lambda}] \]

\[ \Rightarrow \text{Term4} = \frac{\partial^2 N}{\partial \lambda^2} = [-A_4^2 \frac{\partial A_4}{\partial \lambda} - \hat{A}_4 A_4^4 \frac{\partial A_4}{\partial \lambda} + A_4^2 + 2 A_4 A_4^3 \frac{\partial A_4}{\partial \lambda} + A_4^2 \hat{A}_4 A_4 - A_4^2 \hat{A}_4] \]

Note that \( \hat{A}_4 = \frac{\partial A_4}{\partial \lambda} = \frac{1}{\epsilon A^2 + \mu} \Rightarrow \frac{\partial \hat{A}_4}{\partial \lambda} = -\frac{2 \epsilon A^2}{(\epsilon A^2 + \mu)^2} \), \( -\hat{A}_4 \frac{\partial A_4}{\partial \lambda} = -\hat{A}_4^2 = \frac{-1}{(\epsilon A^2 + \mu)^2} \) and \( A_4^2 \hat{A}_4 A_4 = \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2(\epsilon A^2 + \mu)^2} \)

After some algebra:

\[ \text{Term3+Term4} = [A_4^2 \hat{A}_4 A_4 - \hat{A}_4^2] = \left[ \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2(\epsilon A^2 + \mu)^2} - \frac{1}{(\epsilon A^2 + \mu)^2} \right] \]

\[ \Rightarrow [-(\text{Term1+Term1}) + \text{Term2+Term4}] = \left[ \frac{1}{\epsilon A^2 + \mu} - \frac{1}{(\epsilon A^2 + \mu)^2} + \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2(\epsilon A^2 + \mu)^2} - \frac{1}{(\epsilon A^2 + \mu)^2} \right] = \left[ \frac{1}{\epsilon A^2 + \mu} - \frac{1}{(\epsilon A^2 + \mu)^2} + \frac{1}{\epsilon A^2 + \mu} \right] \]

\[ \Rightarrow [-(\text{Term1+Term1}) + \text{Term2+Term4}] = \left[ \frac{1}{\epsilon A^2 + \mu} - \frac{1}{(\epsilon A^2 + \mu)^2} + \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2 A^2(\epsilon A^2 + \mu)} \right] \]

\[ \Rightarrow [-(\text{Term1+Term1}) + \text{Term2+Term4}] = \left[ \frac{1}{\epsilon A^2 + \mu} - \frac{1}{(\epsilon A^2 + \mu)^2} + \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2 A^2(\epsilon A^2 + \mu)} \right] \]

\[ \Rightarrow [-(\text{Term1+Term1}) + \text{Term2+Term4}] = \left[ \frac{-\epsilon + \mu \lambda - \lambda^2}{\sigma^2 A^2(\epsilon A^2 + \mu)} \right] = \left[ \frac{-(\epsilon + \mu \lambda - \lambda^2)}{\sigma^2 A^2(\epsilon A^2 + \mu)} \right] = \frac{1}{\sigma^2} \]
note we use $\Sigma_i^n e_i = 0$

$$\Rightarrow \lim_{\lambda \to 0} \left[ \frac{\partial^2 \ln L}{\partial \mu^2} \right] = -\frac{n}{\sigma^2} - \frac{n}{\sigma^2} = -\frac{2n}{\sigma^2}$$ (66)

End Proof of SOC for $\left[ \frac{\partial^2 \ln L}{\partial \mu^2} \right]$.

$$\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} = \frac{\Sigma_i^n (y - x' \beta + \mu)}{\sigma^4} - n \left[ \frac{\phi(A_1)\left( \frac{\partial A_i^*}{\partial \mu} - A_1^* \frac{\partial A_i}{\partial \mu} A_1 \right)}{[\Phi(A_1) - \Phi(A_2)]} - \frac{\phi(A_2)\left( \frac{\partial A_i^*}{\partial \mu} - A_2^* \frac{\partial A_i}{\partial \mu} A_2 \right)}{[\Phi(A_2)]^2} \right] +$$

$$\Sigma^n \left[ \frac{\phi(A_3)\left( \frac{\partial A_i^*}{\partial \mu} - A_3^* \frac{\partial A_i}{\partial \mu} A_3 \right)}{[\Phi(A_3) - \Phi(A_4)]} - \frac{\phi(A_4)\left( \frac{\partial A_i^*}{\partial \mu} - A_4^* \frac{\partial A_i}{\partial \mu} A_4 \right)}{[\Phi(A_4)]^2} \right] where A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i^*}{\partial \sigma^2}$$

Proof: for Term 1

$$\text{Term 1} = \frac{-\phi(A_2)\left( \frac{\partial A_i^*}{\partial \mu} - A_2^* \frac{\partial A_i}{\partial \mu} A_2 \right)}{[\Phi(A_1) - \Phi(A_2)]}$$

Note that $A_2 \frac{\partial A_2}{\partial \mu} A_2 = \hat{A}_2 A_2^2$

where $\hat{A}_2 = \frac{(\lambda^{-2})_i^2}{2\sigma^2} \Rightarrow N = -\phi_i(A_2)[\hat{A}_2 - \hat{A}_2 A_2^2]$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_2)[-2\hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda} - A_2^2 \frac{\partial \hat{A}_2}{\partial \lambda} + \frac{\partial \hat{A}_2}{\partial \lambda} - \hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda} + \hat{A}_2 A_2^3 \frac{\partial A_2}{\partial \lambda}]$$
\[
\frac{\partial N}{\partial \lambda} = -\phi_i(A_2)[-\hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda} - A_2^2 \frac{\partial A_2}{\partial \lambda} \phi_i(A_2) \frac{\partial \phi_i(A_2)}{\partial A_2} + A_2 \frac{\partial A_2}{\partial \lambda} + \hat{A}_2 A_2^3 \frac{\partial A_2}{\partial \lambda}]
\]

\[
\Rightarrow \text{Term 1} = [-3\hat{A}_2 A_2 - A_2^2 \frac{\partial A_2}{\partial \lambda} + \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 A_2^3]
\]

For Term 2:

\[
\text{Term 2} = -\frac{\partial^2 (A_2) A_2^3}{\phi(A_2) (\Phi(A_1))^2} \frac{\partial A_2}{\partial \mu}, \text{Note that } A_2^2 \frac{\partial A_2}{\partial \mu} = \hat{A}_2 A_2
\]

\[
\Rightarrow N = \phi_i^2(A_2) \hat{A}_2 A_2
\]

\[
\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i^2(A_2) [A_2 \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 \frac{\partial A_2}{\partial \lambda} - 2\hat{A}_2 A_2^2 \frac{\partial A_2}{\partial \lambda}]
\]

\[
\Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi_i^2(A_2) [A_2 \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 \frac{\partial A_2}{\partial \lambda} - 2\hat{A}_2 A_2^2 \frac{\partial A_2}{\partial \lambda}]}{\phi(A_2) \frac{\partial A_2}{\partial \lambda} 2[\Phi(A_3) \frac{\partial (A_4)}{\partial A_2} \frac{\partial (A_4)}{\partial \lambda} - \Phi(A_4) \frac{\partial (A_1)}{\partial \lambda} + \Phi(A_2)]} \Rightarrow N = \phi_i(A_2) [A_2^2 \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 A_2^3]
\]

\[
\phi_i(A_2)[A_2 \hat{A}_{21} + \hat{A}_2^2 - \hat{A}_2 A_2]
\]

\[
\Rightarrow \text{Term 2} = \frac{\partial^2 N}{\partial \lambda^2} = [-A_2 \frac{\partial \hat{A}_{21}}{\partial \lambda} - \hat{A}_{21} - \frac{\partial \hat{A}_2}{\partial \lambda} + 2\hat{A}_2 A_2 + A_2^2 \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_{21} A_2^2 + \frac{\hat{A}_2}{2} A_2 - \hat{A}_2 A_2^3]
\]

Note that \[ \frac{\partial \hat{A}_2}{\partial \lambda} = 0 \]

Also \[ \frac{\partial \hat{A}_{21}}{\partial \lambda} = \frac{-\mu(1+\lambda^{-2})}{4\sigma^2}, \quad \hat{A}_{21} = \frac{-\mu(1+\lambda^{-2})}{4\sigma^2}, \quad \hat{A}_2 A_2 = \frac{\mu(1+\lambda^{-2})}{4\sigma^2}, \] and \[ \hat{A}_2 A_2^2 + 2\hat{A}_2 A_2 = \hat{A}_2 A_2 \]

\[
\Rightarrow \text{Term 1} + \text{Term 2} = [-\hat{A}_{21} - \frac{\partial \hat{A}_2}{\partial \lambda} + \frac{\partial \hat{A}_2}{\partial \lambda}]
\]

After some algebra:

\[
\Rightarrow \text{Term 1} + \text{Term 2} = 0
\]

For Term 3:
Term 3: \[ -\phi(A_4)^4 \left( \frac{\partial A_4}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda} \right) \left( \Phi(A_4) - \Phi(A_4) \right) \]

Note that \(-A_4^2 \frac{\partial A_4}{\partial \lambda} A_4 = -\tilde{A}_4 A_4^2\) where \(\tilde{A}_4 = \frac{1}{2\sigma^2}\)

\[ \Rightarrow N = -\phi_i(A_4) [\tilde{A}_4 - \tilde{A}_4 A_4^2] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_4) [-2\tilde{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^3 \frac{\partial \tilde{A}_4}{\partial \lambda} + \frac{\partial \tilde{A}_4}{\partial \lambda} - \tilde{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} + \tilde{A}_4 A_4^3 \frac{\partial A_4}{\partial \lambda}] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_4) [-2\tilde{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^3 \frac{\partial \tilde{A}_4}{\partial \lambda} + \frac{\partial \tilde{A}_4}{\partial \lambda} - \tilde{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} + \tilde{A}_4 A_4^3 \frac{\partial A_4}{\partial \lambda}] \]

For Term 4:

Term 4: \[ -\phi^2(A_4) A_4^3 \frac{\partial A_4}{\partial \mu} \]

Note that \(A_4^3 \frac{\partial A_4}{\partial \mu} = \tilde{A}_4 A_4 \Rightarrow N = \phi_i^2(A_4) \tilde{A}_4 A_4 \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i^2(A_4) [A_4 \frac{\partial \tilde{A}_4}{\partial \lambda} + \tilde{A}_4 \frac{\partial A_4}{\partial \lambda} - 2\tilde{A}_4 A_4^2 \frac{\partial A_4}{\partial \lambda}] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi_i^2(A_4) [A_4 \frac{\partial \tilde{A}_4}{\partial \lambda} + \tilde{A}_4 \frac{\partial A_4}{\partial \lambda} - 2\tilde{A}_4 A_4^2 \frac{\partial A_4}{\partial \lambda}]}{\phi(A_4) 2\tilde{A}_4 A_4 (\Phi(A_4) + \Phi(A_4))} \]

\[ \Rightarrow N = \phi_i(A_4) [A_4 \frac{\partial \tilde{A}_4}{\partial \lambda} + \frac{\tilde{A}_4}{2} - \tilde{A}_4 A_4^2] = \phi_i(A_4) [A_4 \tilde{A}_4 + \frac{\tilde{A}_4}{2} - \tilde{A}_4 A_4^2] \]

\[ \Rightarrow \text{Term 4} = \frac{\partial^2 N}{\partial \lambda^2} = [-A_4 \frac{\partial \tilde{A}_4}{\partial \lambda} - \tilde{A}_4 + A_4 \frac{\partial \tilde{A}_4}{\partial \lambda} + 2A_4 A_4^2 \frac{\partial A_4}{\partial \lambda} + \tilde{A}_4 A_4^3 + \tilde{A}_4 A_4 - \tilde{A}_4 A_4^2] \]

Note that \(-\frac{\partial \tilde{A}_4}{\partial \lambda} = \frac{1}{2\sigma^2\epsilon \lambda^2 + \mu} \Rightarrow \tilde{A}_4 = \frac{-1}{4\sigma^2\epsilon \lambda^2 + \mu} \Rightarrow \frac{\partial \tilde{A}_4}{\partial \lambda} = \frac{-2\epsilon\lambda}{4\sigma^2\epsilon \lambda^2 + \mu} \)

Also \(-\frac{\partial^2 \tilde{A}_4}{\partial \lambda^2} = \frac{1}{4\sigma^2\epsilon \lambda^2 + \mu} A_4^2 \tilde{A}_4 = \frac{(-\epsilon\lambda^2 - 2\mu\epsilon\lambda - \mu^2)}{4\sigma^2\epsilon \lambda^2 + \mu} \) \(A_4^2 \tilde{A}_4 = \frac{-e^2\lambda^2 - 2\mu\epsilon\lambda - \mu^2}{4\sigma^2\epsilon \lambda^2 + \mu} \), \(A_4^2 \tilde{A}_4 = \frac{-e - \mu\epsilon - \mu^2}{4\sigma^2\epsilon \lambda^2 + \mu} \), \(\frac{\partial \tilde{A}_4}{\partial \lambda} = \frac{-2\epsilon\lambda^2}{4\sigma^2\epsilon \lambda^2 + \mu} \)

After some algebra
Term3 + Term4 = 0, using $\Sigma_{i=1}^{n} e_i = 0$

Term1 + Term2 + Term3 + Term4 = 0

$$\lim_{\lambda \to 0} \left[ \frac{\partial^2 \ln l}{\partial \sigma^2 \partial \mu} \right] = \frac{\Sigma_{i}^n(y - x'\beta + \mu)}{\sigma^4} = 0$$

when we substitute for $\hat{\mu} = 0$ and $\Sigma_{i=1}^{n} e_i = 0$

(67)

End Proof of SOC for $[\frac{\partial^2 \ln l}{\partial \sigma^2 \partial \mu}]$.

$$\frac{\partial^2 \ln L}{\partial \lambda \partial \mu} = -n\left[ \phi(A_1)(\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1) - \phi(A_2)(\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2) \right] -$$

$$\frac{(\phi(A_1)A_1^* - \phi(A_2)A_2^*) (\phi(A_1)\frac{\partial A_1}{\partial \mu} - \phi(A_2)\frac{\partial A_2}{\partial \mu})}{[\Phi(A_1) - \Phi(A_2)]^2} +$$

$$\Sigma_{i}^n \left[ \phi(A_3)(\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3) - \phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4) \right] -$$

$$\frac{\phi(A_3)A_3^* - \phi(A_4)A_4^*) \phi(A_3)\frac{\partial A_3}{\partial \mu} - \phi(A_4)\frac{\partial A_4}{\partial \mu}}{[\Phi(A_3) - \Phi(A_4)]^2}$$

where $A_i^*$, $(i = 1, 2, 3, 4) = \frac{\partial A_i}{\partial A}$

Term1 = $$-\phi(A_2)(\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2)$$

$$\frac{-\phi^2(A_2) A_2^*}{[\Phi(A_1) - \Phi(A_2)]^2}$$

Term2 = $$-\phi^2(A_2) A_2^* \frac{\partial A_2}{\partial \mu}$$

Term3 = $$-\phi(A_4)(\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4)$$

Term4 = $$-\phi^2(A_4) A_4^* \frac{\partial A_4}{\partial \mu}$$

Proof:
For Term 1:

Note that $A_{2} \frac{\partial A_{2}}{\partial \mu} A_{2} = \hat{A}_{2} A_{2}^{2}$ where $\hat{A}_{2} = (\frac{\lambda^{-2}+1}{\sigma})^2 \lambda^{-3}$

Let $N = -\phi_{i}(A_{2})[\hat{A}_{2} - \hat{A}_{2} A_{2}^{2}]$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_{i}(A_{2})[\frac{\partial \hat{A}_{2}}{\partial \lambda} - 2\hat{A}_{2} A_{2} \frac{\partial A_{2}}{\partial \lambda} - A_{2} \frac{\partial \hat{A}_{2}}{\partial \lambda} - \hat{A}_{2} A_{2} \frac{\partial A_{2}}{\partial \lambda} + \hat{A}_{2} A_{2} \frac{\partial A_{2}}{\partial \lambda}]$$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_{i}(A_{2})(\frac{\partial \hat{A}_{2}}{\partial \lambda} - 2\hat{A}_{2} A_{2} \frac{\partial A_{2}}{\partial \lambda} - A_{2} \frac{\partial \hat{A}_{2}}{\partial \lambda})$$

$$\Rightarrow \text{Term1} = [\hat{A}_{2} A_{2} - \hat{A}_{2} A_{2}^{2}]$$

For Term 2:

Note that $A_{2} \frac{\partial A_{2}}{\partial \mu} = \hat{A}_{2} A_{2} \Rightarrow N = \phi_{i}^{2}(A_{2}) \hat{A}_{2} A_{2}$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_{i}^{2}(A_{2})[A_{2} \frac{\partial \hat{A}_{2}}{\partial \lambda} + \hat{A}_{2} \frac{\partial A_{2}}{\partial \lambda} - 2\hat{A}_{2} A_{2} \frac{\partial A_{2}}{\partial \lambda}]$$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_{i}(A_{2})[A_{2} \frac{\partial \hat{A}_{2}}{\partial \lambda} + \hat{A}_{2} \frac{\partial A_{2}}{\partial \lambda}]$$

$$\Rightarrow \text{Term2} = [-A_{2} \frac{\partial \hat{A}_{2}}{\partial \lambda} - \hat{A}_{2} A_{2}^{2} + \hat{A}_{2} A_{2}^{2}]$$

Note that $\frac{\partial \hat{A}_{2}}{\partial \lambda} = \frac{\lambda^{-3} \mu^{2} (\lambda^{-2}+1)^{-1}}{2\mu^{2}} - \frac{3\lambda^{-1}}{2\mu} \Rightarrow \hat{A}_{2} = \frac{\lambda^{-3} \mu^{2} (\lambda^{-2}+1)^{-1}}{2\mu} - \frac{3\lambda^{-1}}{2\mu}$

$$\Rightarrow \frac{\partial \hat{A}_{2}}{\partial \lambda} = \frac{\lambda^{-3} \mu^{2} (\lambda^{-2}+1)^{-1}}{2\mu^{2}} - \frac{3\lambda^{-1}}{2\mu} \Rightarrow \hat{A}_{2} = \frac{\lambda^{-3} \mu^{2} (\lambda^{-2}+1)^{-1}}{2\mu} - \frac{3\lambda^{-1}}{2\mu}$$

Also $\hat{A}_{2} A_{2}^{2} = \frac{\mu \lambda^{-3}}{2\sigma^{2}} + \frac{3\mu \lambda^{-1}(1+\lambda^{-2})}{2\sigma^{2}}$. 

93
\[ -[\text{Term1} + \text{Term2}] = \left[ \frac{\lambda^{-1}}{\mu(\lambda^2 + 1)} - \frac{\mu\lambda^{-3}}{2\sigma^2} - \frac{3\mu\lambda^{-1}}{2\sigma^2} \right] \]

For Term3:

Note that \(-A_4^* \frac{\partial A_4}{\partial \mu} A_4 = -\hat{A}_4 A_4\) where \(\hat{A}_4 = \frac{\left(\mu + \lambda\right)}{\sigma^2}\)

\[ \Rightarrow N = -\phi_i(A_4)[-\hat{A}_4 A_4^2] \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_4) \frac{\partial^2 A_4^*}{\partial \mu \partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} + A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - \frac{\partial A_4^*}{\partial \mu} \frac{\partial A_4}{\partial \lambda} - \hat{A}_4 A_4^2 \frac{\partial A_4}{\partial \lambda} \]

\[ \Rightarrow \text{Term 3} = \hat{A}_4 + A_4 \frac{\partial A_4}{\partial \lambda} - \frac{\partial A_4^*}{\partial \mu} A_4 - \hat{A}_4 A_4^2 \]

For Term4:

Note that \(A_4^* \frac{\partial A_4}{\partial \mu} = -\hat{A}_4 \Rightarrow N = \phi_i^2(A_4)\hat{A}_4\)

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i^2(A_4) \frac{\partial \hat{A}_4}{\partial \lambda} - 2\hat{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} \]

\[ \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{\phi_i^2(A_4) \frac{\partial \hat{A}_4}{\partial \lambda} - 2\hat{A}_4 A_4 \frac{\partial A_4}{\partial \lambda}}{\phi(A_4) \frac{\partial A_4}{\partial \lambda} + \Phi(A_4) \frac{\partial \hat{A}_4}{\partial \lambda} + \Phi(A_4) \frac{\partial A_4}{\partial \lambda} + \Phi(A_4) \frac{\partial \hat{A}_4}{\partial \lambda}} \]

\[ \Rightarrow N = \phi_i(A_4)[A_4 \frac{\partial A_4}{\partial \lambda} - \hat{A}_4 A_4] = \phi_i(A_4)[A_4 \hat{A}_4 - \hat{A}_4 A_4] \]

\[ \Rightarrow \text{Term 4} = \frac{\partial^2 N}{\partial \lambda^2} = \frac{\partial \hat{A}_4}{\partial \lambda} - \hat{A}_4 - A_4 \frac{\partial \hat{A}_4}{\partial \lambda} - \hat{A}_4 A_4^2 \]

Note that \(\frac{\partial A_4}{\partial \lambda} = \frac{-[x + 3\mu \lambda^{-2}]}{\sigma [x \lambda^2 + \mu]} \Rightarrow \hat{A}_{41} = \frac{-[x + 3\mu \lambda^{-2}]}{2\sigma [x \lambda^2 + \mu]}\).
Also \( \frac{\partial^4 A_{11}}{\partial \lambda^4} = \frac{6 \lambda^{-1}}{2(\lambda^2 + \mu)^2} + \frac{2 \lambda^2}{2(\lambda^2 + \mu)^2} + \frac{6 \lambda \mu}{2(\lambda^2 + \mu)^2} \), \( -A_4 \frac{\partial A_4}{\partial \mu} \frac{\partial^2 A_4}{\partial \lambda^2} = -\frac{\varepsilon \lambda^{-1} + \mu \lambda^{-3}}{\sigma^2} \), \( \frac{\partial^2 A_4}{\partial \mu \partial \lambda} = -\frac{2 \lambda^{-1}}{\sigma^2} \), \( A_4 \tilde{A}_{11} = \frac{(\sigma^2 - 2 \varepsilon \lambda^{-1} - 3 \mu^2 \lambda^{-3})}{2 \sigma^2(\varepsilon \lambda^2 + \mu)} \)

\[ \Rightarrow -(\text{Term} 1 + \text{Term} 2) + (\text{Term} 3 + \text{Term} 4) = \frac{(\frac{\varepsilon}{\lambda} + \frac{2 \mu}{\lambda} + 2 \frac{\mu}{\lambda^2})}{2 \sigma^2} = 0 \text{ since } \frac{\varepsilon}{\lambda} = 0 \text{ as } \lambda \to 0, \text{ the numerator goes to zero faster than the denominator.} \]

\[ \Rightarrow -(\text{Term} 1 + \text{Term} 2) + (\text{Term} 3 + \text{Term} 4) = 0 \]

\( \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \lambda \partial \mu} = 0 \) \hspace{1cm} (68)

End Proof of SOC for \( \frac{\partial^2 \ln L}{\partial \lambda \partial \mu} \).

\[
\frac{\partial^2 L}{\partial \sigma^2 \partial \lambda} = -n \left[ \frac{\phi(A_1) (\frac{\partial A_1^*}{\partial \lambda} - A_1 \frac{\partial A_1}{\partial \lambda} A_1) - \phi(A_2) (\frac{\partial A_2^*}{\partial \lambda} - A_2 \frac{\partial A_2}{\partial \lambda} A_2)}{[\Phi(A_1) - \Phi(A_2)]^2} \right] - \\
\left[ \frac{\phi(A_1) A_1^* - \phi(A_2) A_2^*}{[\Phi(A_1) - \Phi(A_2)]^2} \right] + \\
\sum_i \left[ \frac{\phi(A_3) (\frac{\partial A_3^*}{\partial \lambda} - A_3 \frac{\partial A_3}{\partial \lambda} A_3) - \phi(A_4) (\frac{\partial A_4^*}{\partial \lambda} - A_4 \frac{\partial A_4}{\partial \lambda} A_4)}{[\Phi(A_3) - \Phi(A_4)]} \right] - \\
\left[ \frac{\phi(A_3) A_3^* - \phi(A_4) A_4^*}{[\Phi(A_3) - \Phi(A_4)]^2} \right] \\
\text{where } A_i^*, (i = 1, 2, 3, 4) = \frac{\partial A_i}{\partial \sigma^2}
\]

\[ \text{Term} 1 = -\phi(A_2) (\frac{\partial A_2^*}{\partial \lambda} - A_2 \frac{\partial A_2}{\partial \lambda} A_2) \left[ \Phi(A_1) - \Phi(A_2) \right] \]

\[ \text{Term} 2 = -\phi(A_1) A_1^* \frac{\partial A_2}{\partial \lambda} \left[ \Phi(A_1) - \Phi(A_2) \right]^2 \]

\[ \text{Term} 3 = -\phi(A_4) (\frac{\partial A_4^*}{\partial \lambda} - A_4 \frac{\partial A_4}{\partial \lambda} A_4) \left[ \Phi(A_1) - \Phi(A_4) \right] \]

\[ \text{Term} 4 = -\phi(A_1) A_1^* \frac{\partial A_4}{\partial \lambda} \left[ \Phi(A_1) - \Phi(A_4) \right]^2 \]
Proof:

Note that $-A_2^2 \frac{\partial A_2}{\partial x} A_2 = \hat{A}_2 A_2^2$ where $\hat{A}_2 = \frac{(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3}}{2\sigma^2}$

$\Rightarrow N = -\phi_i(A_2)[-\hat{A}_2 + \hat{A}_2 A_2^2]$

$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_2)[2\hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda} + A_2^2 \frac{\partial \hat{A}_2}{\partial \lambda} - \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda} - \hat{A}_2 A_2^2 \frac{\partial A_2}{\partial \lambda}]$

$\frac{\partial N}{\partial x} = -\phi_i(A_2)[2\hat{A}_2 A_2 \frac{\partial A_2}{\partial x} + A_2^2 \frac{\partial \hat{A}_2}{\partial x} - \frac{\partial \hat{A}_2}{\partial x} + \hat{A}_2 A_2 \frac{\partial A_2}{\partial x} - \hat{A}_2 A_2^2 \frac{\partial A_2}{\partial x}]$

Term1 = $[3\hat{A}_2 A_2 + A_2^2 \frac{\partial \hat{A}_2}{\partial x} - \frac{\partial \hat{A}_2}{\partial x} - \hat{A}_2 A_2^2]$

For Term 2:

Note that $-A_2^2 \frac{\partial A_2}{\partial x} = \hat{A}_2 A_2 \Rightarrow N = \phi_i^2(A_2)\hat{A}_2 A_2$

$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i^2(A_2)[A_2 \frac{\partial \hat{A}_2}{\partial \lambda} + \hat{A}_2 \frac{\partial A_2}{\partial \lambda} - 2\hat{A}_2 A_2 \frac{\partial A_2}{\partial \lambda}]$

$\Rightarrow \frac{\partial N}{\partial x} = \lambda^{-3}(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3} \frac{\partial \hat{A}_2}{\partial x} + \hat{A}_2 \frac{\partial A_2}{\partial x} - A_2^2 \frac{\partial \hat{A}_2}{\partial x} - \hat{A}_2 A_2 \frac{\partial A_2}{\partial x} + \hat{A}_2 A_2^2$

note that $\frac{\partial \hat{A}_2}{\partial x} = \frac{\lambda^{-3}(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3}}{2\sigma^2}$

$\Rightarrow \frac{\partial \hat{A}_2}{\partial \lambda} = \frac{\lambda^{-3}(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3}}{2\sigma^2}$

$A_2 \frac{\partial \hat{A}_2}{\partial \lambda} = \frac{\lambda^{-3}(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3}}{2\sigma^2} A_2 \frac{\partial \hat{A}_2}{\partial \lambda} = \frac{\lambda^{-3}(\lambda^{-2}+1)^{-\frac{1}{2}} \lambda^{-3}}{2\sigma^2}$

Term2 = $[A_2 \frac{\partial \hat{A}_2}{\partial x} + \hat{A}_2 + 2\hat{A}_2 A_2 \frac{\partial \hat{A}_2}{\partial x} - \hat{A}_2 A_2 \frac{\partial \hat{A}_2}{\partial x} - \hat{A}_2 A_2^2 - \hat{A}_2 A_2^2]$

Also $-\hat{A}_2 A_2^2 = \frac{\lambda^2 \lambda^{-3}}{2\sigma^2} + \frac{3\mu^2 \lambda^{-1}}{4\sigma^2}$,
For Term 3:

Note that $A_3 \frac{\partial A_4}{\partial \lambda} A_4 = -\hat{A}_4 A_4^3$ where $\hat{A}_4 = \frac{\varepsilon + \mu \lambda^{-1}}{2\sigma^2}$

$$\Rightarrow N = -\phi_i(A_4)[-\hat{A}_4 + \hat{A}_4 A_4^3]$$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi_i(A_4)[2\hat{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} + A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} + \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 A_4 \frac{\partial A_4}{\partial \lambda} - \hat{A}_4 A_4^3 \frac{\partial A_4}{\partial \lambda}]$$

$$\frac{\partial N}{\partial \lambda} = -\phi_i(A_4)\frac{\partial A_4}{\partial \lambda} + \frac{\partial \hat{A}_4}{\partial \lambda}$$

Term3$=\{3\hat{A}_4 A_4 + A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} - \frac{\partial \hat{A}_4}{\partial \lambda} - \hat{A}_4 A_4^3]\$

For Term 4:

Note that $A_3 \frac{\partial A_4}{\partial \lambda} A_4 = -\hat{A}_4 A_4^3$ \Rightarrow $N = \phi_i(A_4)\hat{A}_4 A_4$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i(A_4)[A_4 \frac{\partial A_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - 2\hat{A}_4 A_4^2 \frac{\partial A_4}{\partial \lambda}]$$

$$\Rightarrow \frac{\partial N}{\partial \lambda} = \phi_i(A_4)\frac{\partial A_4}{\partial \lambda} - 2\hat{A}_4 A_4^2 \frac{\partial A_4}{\partial \lambda}$$

$$\Rightarrow \frac{\partial^2 N}{\partial \lambda^2} = \{A_4 \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 \frac{\partial \hat{A}_4}{\partial \lambda} - 2\hat{A}_4 A_4 - \hat{A}_4 A_4^2 - 2\hat{A}_4 A_4 + \hat{A}_4 A_4^3\}$$

$$\Rightarrow \text{Term4} = \{A_4 \frac{\partial \hat{A}_4}{\partial \lambda} + \hat{A}_4 + A_4^2 \frac{\partial A_4}{\partial \lambda} - 2\hat{A}_4 A_4 - \hat{A}_4 A_4^2 - \hat{A}_4 A_4^3\}$$

Note that $-\frac{\partial A_4}{\partial \lambda} = \frac{-\mu \lambda^{-1}}{\sigma^2 |\lambda^2 + \mu|}$ \Rightarrow $\hat{A}_4 = \frac{-\mu \lambda^{-1}}{2\sigma^2 |\lambda^2 + \mu|}$ \Rightarrow $\frac{\partial \hat{A}_4}{\partial \lambda} = \frac{-\mu \lambda^{-2}}{2\sigma^2 |\lambda^2 + \mu|^2} + \frac{2\mu}{2\sigma^2 |\lambda^2 + \mu|^2}$
Also \( \frac{\partial A_4}{\partial \sigma} = \frac{\mu \lambda - u^2 \lambda^3}{2\sigma^2[\lambda^2 + \mu]^3} \),
\( A_4 \frac{\partial A_4}{\partial \sigma} = \frac{\mu \lambda - u^2 \lambda^3}{2\sigma^2[\lambda^2 + \mu]^3} \),
\( \frac{\partial A_4}{\partial \lambda} \frac{\partial A_4}{\partial \lambda} \frac{\partial A_4}{\partial \lambda} \frac{\partial A_4}{\partial \lambda} \)

\(-\hat{A}_4 A_4^2 = \frac{\mu \lambda - u^2 \lambda^3 + \mu \lambda^3}{2\sigma^2[\lambda^2 + \mu]^3} \)

\(-[\text{Term1+Term2}] + [\text{Term3+Term4}] = \left[ \frac{\lambda - 1}{2\sigma^4[\lambda^2 + 1]} - \frac{\mu \lambda - u^2 \lambda^3 - 3\mu^2 \lambda^3 - 2\sigma^2[\lambda^2 + \mu]}{4\sigma^4[\lambda^2 + \mu]^2} - \frac{\lambda - 1}{2\sigma^2[\lambda^2 + \mu]^3} \right]
\Rightarrow \left[ \frac{-3\mu^2 \lambda - 1}{4\sigma^4} + \frac{-\mu^2 \lambda - 1}{2\sigma^2[\lambda^2 + \mu]^2} + \frac{\lambda - 1}{2\sigma^2[\lambda^2 + \mu]^3} \right] = 0 \text{ since } \mu = 0 \text{ as } \lambda \to 0.

\(-[\text{Term1+Term2}] + [\text{Term3+Term4}] = 0 \text{ so } \lim_{\lambda \to 0} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = 0 \quad (69)

End Proof of SOC for \( \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = 0 \).
Note that $-A_2^* \frac{\partial A_2}{\partial \sigma} A_2 = -A_2 A_2^*$, $N = -\phi(A_2) [-A_2 A_2^* + \frac{\partial A_2^*}{\partial \sigma}] \Rightarrow \frac{\partial N}{\partial \lambda} = -\phi(A_2) [-2A_2 A_2^* + \frac{\partial A_2^*}{\partial \lambda} - A_2^2 \frac{\partial A_2}{\partial \lambda}] + \\
\frac{\partial^2 A_2^*}{\partial \sigma^2} \frac{\partial A_2}{\partial \lambda} + A_2^2 A_2^* \frac{\partial A_2}{\partial \lambda} - A_2 \frac{\partial A_2^*}{\partial \sigma} A_2^2$]
\[
\frac{\partial D}{\partial \lambda} = \frac{1}{\phi(A_2)} \frac{\partial}{\partial \lambda} \frac{\phi(A_1) \frac{\partial A_1}{\partial \lambda} - \phi(A_2) \frac{\partial A_2}{\partial \lambda} - 1} \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{-2A_2 A_2^* \frac{\partial A_2^*}{\partial \lambda} - A_2^2 + \frac{\partial^2 A_2^*}{\partial \sigma^2} \frac{\partial A_2}{\partial \lambda} - A_2^2}{\frac{\phi(A_1) \frac{\partial A_1}{\partial \lambda} - \phi(A_2) \frac{\partial A_2}{\partial \lambda}}{\phi(A_2) \frac{\partial A_2}{\partial \lambda} - 1}} = [-2A_2 A_2^* \frac{\partial A_2^*}{\partial \lambda} - A_2^2 + \\
\frac{\partial^2 A_2^*}{\partial \sigma^2} + A_2^2 A_2^* - A_2 \frac{\partial A_2^*}{\partial \sigma^2}]$

Term 1 = $[-2A_2 A_2^* \frac{\partial A_2^*}{\partial \lambda} - A_2^2 + \frac{\partial^2 A_2^*}{\partial \sigma^2} \frac{\partial A_2}{\partial \lambda} + A_2^2 A_2^* - A_2 \frac{\partial A_2^*}{\partial \sigma^2}]$

For Term 2:

Note that $N = [\phi(A_1) A_1^* - \phi(A_2) A_2^*]^2 = -\phi^2(A_2) A_2^2$

$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi^2(A_2) [-A_2 \frac{\partial^2 A_2}{\partial \lambda} + A_2^2 \frac{\partial A_2}{\partial \sigma^2}]$ and

$D = [\Phi(A_2) - \Phi(A_4)]^2 \Rightarrow \frac{\partial D}{\partial \lambda} = 2 \Phi(A_2) \phi(A_1) \frac{\partial A_2}{\partial \lambda} - 2 \Phi(A_2) \phi(A_1) \frac{\partial A_2}{\partial \lambda} + \Phi(A_1) \phi(A_2) \frac{\partial A_2}{\partial \lambda} + 2 \Phi(A_2) \phi(A_2) \frac{\partial A_2}{\partial \lambda}$

let $\hat{A}_2 = \frac{\partial A_2}{\partial \lambda} \Rightarrow \frac{\partial N}{\partial \lambda} = \frac{-\phi(A_2) [-A_2 \frac{\partial A_2}{\partial \lambda} + A_2 \frac{\partial A_2}{\partial \sigma^2}] + \frac{\partial A_2^*}{\partial \lambda}}{\Phi(A_1) \frac{\partial A_2}{\partial \lambda} - \Phi(A_2) \frac{\partial A_2}{\partial \lambda} + \Phi(A_1) \hat{A}_2} = 0$

$= 0$

Term 2 = $[-A_2^* \frac{\partial A_2}{\partial \lambda} - \hat{A}_2 \frac{\partial A_2^*}{\partial \lambda} + 2 A_2 A_2^* \frac{\partial A_2}{\partial \lambda} + A_2^2 + A_2^* \hat{A}_2 A_2 - A_2^2 A_2^*]$}

Note that $\hat{A}_2 = \frac{-1}{\hat{A}_2} \Rightarrow \frac{\partial \hat{A}_2}{\partial \lambda} = 0$, $A_2^* \hat{A}_2 A_2 = A_2^2$, $\frac{\partial^2 A_2^*}{\partial \lambda^2} = \frac{3}{\pi^2}, -A_2 \frac{\partial A_2^*}{\partial \lambda} = -3 A_2^2, -\hat{A}_2 \frac{\partial A_2^*}{\partial \lambda} = -\frac{1}{\hat{A}_2^2}$

$N = \frac{1}{\hat{A}_2^2} - 2 A_2^2$

For Term 3:

$\Rightarrow \frac{\partial N}{\partial \lambda} = -\phi(A_2) [-2A_4 A_4 \frac{\partial A_4^*}{\partial \lambda} - A_4^2 \frac{\partial A_4}{\partial \lambda} + \frac{\partial^2 A_4^*}{\partial \sigma^2} \frac{\partial A_4}{\partial \lambda} + A_4^2 \frac{\partial A_4}{\partial \lambda} - A_4 \frac{\partial A_4^*}{\partial \sigma^2}]$
\[
\frac{\partial^N}{\partial x} = [-2A_4A_4 \frac{\partial A_4}{\partial x} - A_4^2 + \frac{\partial^2 A_4}{\partial x^2} + A_4^2 A_4^2 - A_4 \frac{\partial A_4}{\partial x}] \\
\]

Term 3 = \([-2A_4A_4 \frac{\partial A_4}{\partial x} - A_4^2 + \frac{\partial^2 A_4}{\partial x^2} + A_4^2 A_4^2 - A_4 \frac{\partial A_4}{\partial x}]\]

For Term 4:
\[
\frac{\partial^N}{\partial x} = [-A_4 \frac{\partial A_4}{\partial x} - \hat{A}_4 \frac{\partial A_4}{\partial x} + 2A_4A_4 \frac{\partial A_4}{\partial x} + A_4^2 + \hat{A}_4 A_4 - A_4^2 A_4^2] \\
\]

Term 4 = \([-A_4 \frac{\partial A_4}{\partial x} - \hat{A}_4 \frac{\partial A_4}{\partial x} + 2A_4A_4 \frac{\partial A_4}{\partial x} + A_4^2 + \hat{A}_4 A_4 - A_4^2 A_4^2]\]

Note that: \(\hat{A}_4 = -\frac{1}{\sigma^4} \Rightarrow \frac{\partial \hat{A}_4}{\partial x} = 0\), \(A_4 \hat{A}_4 A_4 = A_4^2\), \(\frac{\partial^2 \hat{A}_4}{\partial x^2} = \frac{3}{\sigma^4}\), \(-A_4 \frac{\partial A_4}{\partial x^2} = -3A_4^2\), \(-\hat{A}_4 \frac{\partial A_4}{\partial x} = -\frac{1}{\sigma^4}\)

\(
\text{Term 3 + Term 4} = \left[ \frac{1}{2\sigma^4} - 2A_4^2 \right]
\)

\(-\text{Term 1 + Term 2} + \text{Term 3 + Term 4} = 2A_4^2 - 2A_4^2 = \frac{2}{\sigma^4} \left[ \mu^2 \lambda^2 + \mu^2 \lambda^2 - \varepsilon^2 \lambda^2 + 2 \varepsilon \mu - \mu^2 \lambda^2 \right] = \frac{\mu^2 + 2 \varepsilon \mu}{\lambda^2} \frac{2 \varepsilon \mu}{\lambda^2} \frac{2 \varepsilon \mu}{\lambda^2} \frac{2 \varepsilon \mu}{\lambda^2} \frac{2 \varepsilon \mu}{\lambda^2} \frac{2 \varepsilon \mu}{\lambda^2}
\)

\[
\lim_{\lambda \rightarrow 0} \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{\sum_i^n (y - x' \beta + \mu)^2}{\sigma^6} + \frac{\sum_i^n (y - x' \beta + \mu)^2}{2\sigma^6} = \frac{n}{2\sigma^4} - \frac{\sum_i^n (y - x' \beta)^2}{\sigma^6} = \frac{n}{2\sigma^4} (70)
\]

End Proof of \(\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\).
\[
\frac{\partial^2 \ln L}{\partial \lambda^2} = -n \left[ \frac{\phi(A_1)(\frac{\partial A_1}{\partial \lambda}) - A_1^2 A_1 \frac{\partial A_1}{\partial \lambda} - \phi(A_2)(\frac{\partial A_2}{\partial \lambda}) - A_2^2 A_2 \frac{\partial A_2}{\partial \lambda}}{[\Phi(A_1) - \Phi(A_2)]} - \frac{(\phi(A_1)A_1^* - \phi(A_2)A_2^*)^2}{[\Phi(A_1) - \Phi(A_2)]^2} \right] + \\
\sum^n_i \left[ \frac{\phi(A_3)(\frac{\partial A_3}{\partial \lambda}) - A_3^2 A_3 \frac{\partial A_3}{\partial \lambda} - \phi(A_4)(\frac{\partial A_4}{\partial \lambda}) - A_4^2 A_4 \frac{\partial A_4}{\partial \lambda}}{[\Phi(A_3) - \Phi(A_4)]} - \frac{(\phi(A_3)A_3^* - \phi(A_4)A_4^*)^2}{[\Phi(A_3) - \Phi(A_4)]^2} \right] \\
\text{where } A_i^*(i = 1, ..., 4) = \frac{\partial A_i(i = 1, ..., 4)}{\partial \lambda}
\]

**Proof**

For Term 1:

Note that 
\(-A_2 \frac{\partial A_2}{\partial \lambda} A_2 = -A_2 A_2^* N = -\phi(A_2)[A_2 A_2^* + \frac{\partial A_2}{\partial \lambda}] \Rightarrow \frac{\partial A}{\partial \lambda} = -\phi(A_2)\left[ \frac{\partial A_2}{\partial \lambda} - 2A_2 A_2^* \frac{\partial A_2}{\partial \lambda} \right] - \\
\frac{A_2^2 \frac{\partial A_2}{\partial \lambda} - \frac{\partial A_2}{\partial \lambda} \frac{\partial A_2}{\partial \lambda} A_2 + A_2^2 A_2^2 \frac{\partial A_2}{\partial \lambda}}{A_2^2 \frac{\partial A_2}{\partial \lambda} - A_2^2 A_2^2 + A_2^2 A_2^2} \right] \text{ where } A_2^2 = \frac{\partial A_2}{\partial \lambda}

\[
\text{Term1}=\frac{\partial^2 A_2^*}{\partial \lambda^2} - 3A_2 \frac{\partial A_2^*}{\partial \lambda} - A_2^* A_2^* + A_2 A_2^2
\]

For Term 2:

Note that 
\(N = \phi^2(A_2)A_2^2 \Rightarrow \frac{\partial N}{\partial \lambda} = \phi^2(A_2)2[A_2 A_2^3 + A_2 \frac{\partial A_2}{\partial \lambda}] \text{ and } \\
D = [\Phi(A_3) - \Phi(A_4)]^2
\[ \frac{\partial^2 N}{\partial \lambda^2} = 2\Phi(A_1)\phi(A_1)\frac{\partial A_1}{\partial \lambda} - 2[\Phi(A_2)\phi(A_1)\frac{\partial A_1}{\partial \lambda} + \Phi(A_1)\phi(A_2)\frac{\partial A_2}{\partial \lambda}] + 2\Phi(A_2)\phi(A_2)\frac{\partial A_2}{\partial \lambda} \]

let \( N = \phi(A_2)[-A_2 A_2^2 + \frac{\partial A_2}{\partial \lambda}] \)

\[ \frac{\partial^2 N}{\partial \lambda^2} = \left[ -\frac{\partial^2 A_2}{\partial \lambda^2} + A_2^2 + 2A_2 \frac{\partial A_2}{\partial \lambda} + A_2 \frac{\partial A_2}{\partial \lambda} - A_2^2 A_2^2 \right] \]

Term 2 = \[ -\frac{\partial^2 A_2}{\partial \lambda^2} + A_2^2 + 2A_2 \frac{\partial A_2}{\partial \lambda} + A_2 \frac{\partial A_2}{\partial \lambda} - A_2^2 A_2^2 \]

\[ \Rightarrow [\text{Term 1} + \text{Term 2}] = 0 \]

For Term 3:

\[ \frac{\partial N}{\partial \lambda} = -\phi(A_4)[\frac{\partial^2 A_4}{\partial \lambda^2} - A_4^2 A_4 - 2A_4 A_4 \frac{\partial A_4}{\partial \lambda} - A_4^2 A_4 - A_4 \frac{\partial A_4}{\partial \lambda} A_4 + A_4^2 A_4^2] \]

\[ \frac{\partial N}{\partial \lambda} = \left[ \frac{\partial^2 A_4}{\partial \lambda^2} - 3A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 - A_4^2 + A_4^2 A_4^2 \right] \]

Term 3 = \[ \frac{\partial^2 A_4}{\partial \lambda^2} - 3A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 - A_4^2 + A_4^2 A_4^2 \]

For Term 4:

\[ \frac{\partial^2 N}{\partial \lambda^2} = \left[ - \frac{\partial^2 A_4}{\partial \lambda^2} + 2A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 + A_4^2 + A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 - A_4^2 A_4^2 \right] \]

Term 4 = \[ - \frac{\partial^2 A_4}{\partial \lambda^2} + 2A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 + A_4^2 + A_4 A_4 \frac{\partial A_4}{\partial \lambda} A_4 - A_4^2 A_4^2 \]

\[ \text{Term 3} + \text{Term 4} = 0 \]

End Proof \[ \left[ \frac{\partial^2 \ln L}{\partial \lambda^2} \right]. \]

The SOC for the N-TN is evaluated by first allow \( B \to \infty \) and then allow \( \lambda \to 0 \). The results are the same except there is no upper bound. For the N-HN model we allow \( B \to \infty \) then set \( \mu = 0 \), then allow
D Stable stationary point for the N–DTN, N–TN & N-HN

The Hessian for $(\beta, \sigma, \mu, B, \lambda)$ evaluated at $\theta^* = (\beta, \sigma^2, \mu = 0, \lambda = 0)$.

\[
H(\theta^*) = \begin{bmatrix}
-\frac{\sum_{i=1}^n x_i x_i'}{\sigma_v^2} & 0 & \frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 & 0 \\
0 & \frac{-n}{2\sigma_v^2} & 0 & 0 & 0 \\
\frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 & -\frac{2n}{\sigma_v^2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

$H(\theta^*)$ is negative semidefinite with two zero eigenvalues. The eigenvectors associated with zero eigenvalues are:

\[
H(\theta^*)V = \begin{bmatrix}
-\frac{\sum_{i=1}^n x_i x_i'}{\sigma_v^2} & 0 & \frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 & 0 \\
0 & \frac{-n}{2\sigma_v^2} & 0 & 0 & 0 \\
\frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 & -\frac{2n}{\sigma_v^2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
v_k \\
v_{k+1} \\
v_{k+2} \\
v_{k+3} \\
v_{k+4}
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

solve for the eigenvectors associated with the first zero eigenvalue:
\[
\frac{-\Sigma_{i=1}^{n} x_i x_i'}{\sigma_v^2} v_k + \frac{\Sigma_{i=1}^{n} x_i}{\sigma_v^2} v_{k+2} = 0 \Rightarrow v_k = 0 \text{ and } v_{k+2} = 0
\] (71)

\[
\frac{-n}{2\sigma_v^2} v_{k+1} = 0 \Rightarrow v_{k+1} = 0
\] (72)

\[
\frac{\Sigma_{i=1}^{n} x_i}{\sigma_v^2} v_k + \frac{-2n}{\sigma_v^2} v_{k+2} = 0 \Rightarrow v_k = 0 \text{ and } v_{k+2} = 0
\] (73)

\[
0v_k + 0v_{k+2} + 0v_{k+3} + 0v_{k+4} = 0 \Rightarrow v_{k+3} = 1
\] (74)

\[
0v_k + 0v_{k+2} + 0v_{k+3} + 0v_{k+4} = 0 \Rightarrow v_{k+4} = 1
\] (75)

solve for the eigenvectors associated with the second zero eigenvalue:

\[
\frac{-\Sigma_{i=1}^{n} x_i x_i'}{\sigma_v^2} v_k + \frac{\Sigma_{i=1}^{n} x_i}{\sigma_v^2} v_{k+2} = 0 \Rightarrow v_k = 0 \text{ and } v_{k+2} = 0
\] (76)

\[
\frac{-n}{2\sigma_v^2} v_{k+1} = 0 \Rightarrow v_{k+1} = 0
\] (77)

\[
\frac{\Sigma_{i=1}^{n} x_i}{\sigma_v^2} v_k - \frac{2n}{\sigma_v^2} v_{k+2} = 0 \Rightarrow v_k = 0 \text{ and } v_{k+2} = 0
\] (78)

\[
0v_k + 0v_{k+2} + 0v_{k+3} + 0v_{k+4} = 0 \Rightarrow v_{k+3} = 1
\] (79)

\[
0v_k + 0v_{k+2} + 0v_{k+3} + 0v_{k+4} = 0 \Rightarrow v_{k+4} = 1
\] (80)

Let \(\gamma_1\) and \(\gamma_2\) be some very small positive numbers and restrict the eigenvectors to be positive because \(\lambda > 0\).

\[
\begin{bmatrix}
0 \\
0 \\
1 \\
0
\end{bmatrix}
\]

The eigenvectors are \(z = z_1 + z_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\) and \(z = z_1 \gamma_1 + z_2 \gamma_2 = \begin{bmatrix} 0 \\ 0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix}\). The change in the
loglikelihood will be evaluated based on the number of non–zero elements in \( z \), here there are two non–zero elements, \( (\lambda \text{ and } B) \), see the footnote in Waldman(1982).

\[
\text{N-DTN} : \quad \Delta \ln l = \ln l(\theta^* + z\gamma) - \ln l(\theta^*) = -n[\ln(\Phi(\frac{\gamma_1(\gamma_2^{-2} + 1)^{\frac{1}{2}}}{\sigma}) - \Phi(0))] + \\
\sum_i^n [\ln(\Phi(\frac{(\gamma_1 + e_i)\gamma_2 + \gamma_1^{-1}}{\sigma})) - \Phi(\frac{e_i\gamma_2}{\sigma})]^{20}
\]

(81)

The paper uses a third order Taylor series expansion around the points \((\hat{\lambda} = 0, B \to \infty)\) and use the fact \( \frac{\partial}{\partial \lambda} \hat{\lambda} = 0 \) to recover how these parameters influence the behaviour of the loglikelihood when \( \hat{\sigma}_u = 0 \). Recall that \( \lim_{\lambda \to 0} \frac{\partial \ln l}{\partial B} = 0 \) therefore we can choose any value for \( B \), here I choose \( B \to \infty \), this is because in the limit one is unable to discriminate among all the models above that is, \( N - DTN = N - TN = N - HN \). The paper perturbs the loglikelihood of the N–DTN model and examine if OLS is a stable solution in the neighborhood of small \( \sigma_u \). If \( \Delta \ln l \leq 0 \) OLS yields a higher value of the loglikelihood and \( u_i = 0 \ (i = 1, ... n) \) and if \( \Delta LnL > 0 \) then MLE yields a higher value of the loglikelihood and \( u_i \geq 0 \) for \( (i = 1, ... n) \).

\[
\Delta \ln l \approx \Delta \ln l(\hat{\lambda} = 0, B \to \infty) + \frac{\partial}{\partial B} \Delta \ln l(\gamma_1) + \frac{\partial}{\partial \lambda} \Delta \ln l(\gamma_2) + \frac{1}{2} \frac{\partial^2}{\partial B^2} \Delta \ln l(\gamma_1)^2 + \frac{\partial^2}{\partial \lambda^2} \Delta \ln l(\gamma_2)^2 \\
+ 2 \frac{\partial^2}{\partial B \partial \lambda} \Delta \ln l(\gamma_1)(\gamma_2) + \frac{1}{6} \frac{\partial^3}{\partial B^3} \Delta \ln l(\gamma_1)^2 + 3 \frac{\partial^3}{\partial B^2 \partial \lambda} \Delta \ln l(\gamma_1)^2(\gamma_2) + 3 \frac{\partial^3}{\partial B^2 \lambda^2} \Delta \ln l(\gamma_1)(\gamma_2)^2 \\
+ \frac{\partial^3}{\partial \lambda^3} \Delta \ln l(\gamma_2)^3) + o(\gamma_2^3)
\]

Note that any terms multiply by \( B \) will go to zero as \( B \to \infty \), this is because the exponent dominates.

\[
\implies \ln L(\hat{\lambda} = 0, B \to \infty) = \frac{\partial^2 \Delta \ln l}{\partial B^2} = \frac{\partial^2 \Delta \ln l}{\partial B \partial \lambda} = \frac{\partial^3 \Delta \ln l}{\partial B^2 \partial \lambda} = \frac{\partial^3 \Delta \ln l}{\partial B \partial \lambda^2} = 0
\]

Only the partial derivatives terms associated with lambda need to be evaluated.

\[\text{21}\] The error rate for the remainder term is associated with only the difference between the estimate of lambda and the true lambda value.
1. The paper uses the fact that $\frac{\sum_{i=1}^{n} e_i}{n} = 0$ (the sum of OLS residuals equal to zero because $E(\varepsilon) = -E(u) = 0$ when $\lambda = \hat{\lambda} = 0$)

2. $\frac{\partial \ln L}{\partial \lambda} = 0$ coming from the FOC

3. Also $\frac{\partial^2 \ln L}{\partial \lambda^2} \gamma_2 = 0$ (coming from the fact that this is derived above by using the zero eigenvalue/vector that is $H(\theta^*)V = 0$).

The only important term that needs to be evaluated is $\frac{\partial^3 \ln L}{\partial \lambda^3}$.

$$\frac{\partial \ln L}{\partial \lambda} = \sum_i \left[ \frac{\phi(A)^3}{\nu} \right] = 0 \text{ since the } \frac{1}{\nu} \sum e_i = 0 \text{ where } A = \frac{\hat{\lambda}}{\lambda} \text{ note that } A = 0 \text{ for } \hat{\lambda} = 0$$

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \sum_i \left[ \frac{\phi(A)A e_i^2}{[1-\Phi(A)]^2} - \frac{\phi(A) e_i^2}{[1-\Phi(A)]^2} \right] = \sum_i \left[ -\frac{\phi(A) e_i^2}{[1-\Phi(A)]^2} \right] = 0 \text{ see 3 above}$$

Let Term 1 = $\frac{\phi(A)A e_i^2}{[1-\Phi(A)]^2}$ and Term 2 = $\frac{\phi(A) e_i^2}{[1-\Phi(A)]^2}$. For Term 1

$$\frac{\partial^3 \ln L}{\partial \lambda^3} = \sum_i \left[ \frac{\phi(A) A e_i^2}{[1-\Phi(A)]^3} \right] = \frac{\phi(A) A^3}{[1-\Phi(A)]^3}$$

is the only relevant term, since $A = 0$ for $\lambda \to 0$

For Term 2

$$\frac{\partial^3 \ln L}{\partial \lambda^3} = -\frac{1-\Phi(A)^2}{1-\Phi(A)^3} \frac{2\phi(A) A^2 e_i^2}{[1-\Phi(A)]^3} + \frac{\phi(A) A e_i^2}{[1-\Phi(A)]^3} - \frac{\phi(A) A^2 e_i^2}{[1-\Phi(A)]^3} (2-\Phi(A)) \phi(A) \frac{\partial^2 A}{\partial \lambda^2} = \frac{-\phi^3(1-\Phi(A)) e_i^2}{[1-\Phi(A)]^3}$$

$-2\phi^3(1-\Phi(A)) e_i^2$ the only relevant term, since $A = 0$ for $\lambda \to 0$

$$\Rightarrow \text{ Term 1 + Term 2 = } \sum_i \left[ \frac{\phi(A) A^3}{[1-\Phi(A)]^3} - \frac{2\phi^3(1-\Phi(A)) e_i^2}{[1-\Phi(A)]^3} \right] = \frac{\sum e_i^3}{\sigma^3} \left[ \frac{2}{\sqrt{2\pi}} \left( \frac{\pi - 4}{\pi} \right) \gamma_2 \right]$$

For the N-DTN:

$$\Delta LnL = \frac{1}{6} \frac{\sum e_i^3}{\sigma^3} \frac{2}{\sqrt{2\pi}} \left( \frac{\pi - 4}{\pi} \right) \gamma_2 + o(\gamma_2)$$

Note that $\gamma_2 = \gamma$ as in the case of the N-HN model.

**D.0.5 N-TN Model**

The Hessian evaluate at $\theta^* = (\hat{\beta}, \hat{\sigma^2}, \hat{\mu} = 0, \hat{\lambda} = 0)$ is:
\[
H(\theta^*) = \begin{bmatrix}
-\frac{\sum_{i=1}^n x_i x_i'}{\sigma_v^2} & 0 & \frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 \\
0 & -\frac{n}{2\pi} & 0 & 0 \\
\frac{\sum_{i=1}^n x_i}{\sigma_v^2} & 0 & -\frac{2n}{\sigma_v^2} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

There is one zero eigenvalue. The change in the loglikelihood is:

For the N-TN:  \[ \Delta \ln l = \ln l(\theta^* + z\gamma) - \ln l(\theta^*) = -n[\ln(1 - \Phi(0)) + \sum_i^n \ln(1 - \Phi(\frac{e_i\gamma_2}{\sigma}))] \] (83)

Using a third order Taylor series expansion around the points \( \hat{\lambda} = 0 \)

\[ \Delta \ln l = \Delta \ln l(\hat{\lambda} = 0) + \frac{\partial \ln l}{\partial \lambda}(\gamma_2) + \frac{1}{2} \frac{\partial^2 \ln l}{\partial \lambda^2}(\gamma_2)^2 + \frac{1}{6} \frac{\partial^3 \ln l}{\partial \lambda^3}(\gamma_2)^3 + o(\gamma_2^4) \]

For the N-TN:  \[ \Delta LnL = \frac{1}{6} \frac{\sum_{i}^3}{\sigma_v^3} \frac{2}{\sqrt{2\pi}} \frac{|\pi - 4|}{\pi} \gamma_2^3 + o(\gamma_2^4) \] (84)

Note that \( \gamma_2 = \gamma \) as in the case of the N-HN model.

D.0.6 N-HN

The Hessian evaluate at \( \theta^* = (\hat{\beta}, \sigma^2, \hat{\lambda} = 0) \) is:

\[
H(\theta^*) = \begin{bmatrix}
-\frac{\sum_{i=1}^n x_i x_i'}{\sigma_v^2} & 0 & \frac{2}{\sqrt{2\pi}} \frac{1}{\sigma_v^2} \sum_{i=1}^n x_i & 0 \\
0 & -\frac{n}{2\pi} & 0 & 0 \\
\frac{2}{\sqrt{2\pi}} \frac{1}{\sigma_v^2} \sum_{i=1}^n x_i & 0 & -\frac{2n}{\sigma_v^2} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

The eigenvector:

\[
H(\theta^*)V = \begin{bmatrix}
-\frac{\sum_{i=1}^n x_i x_i'}{\sigma_v^2} & 0 & \frac{2}{\sqrt{2\pi}} \frac{1}{\sigma_v} \sum_{i=1}^n x_i & 0 \\
0 & -\frac{n}{2\pi} & 0 & 0 \\
\frac{2}{\sqrt{2\pi}} \frac{1}{\sigma_v^2} \sum_{i=1}^n x_i & 0 & -\frac{2n}{\sigma_v^2} & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
v_k \\
v_{k+1} \\
v_{k+2} \\
0
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]
There is one zero eigenvalue. Note the first row $H(\theta^*) = \left[ \begin{array}{cc} -\frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} & 0 \\ \frac{2}{\sqrt{2\pi}} \sigma \end{array} \right]$. If the first row is multiply by $\frac{2}{\sqrt{2\pi}} \sigma$ and add it to the second row it will produce a zero row, similarly we will get a zero column. Since there is one zero row and one zero column $H(\theta^*)$ is singular with one zero eigenvalue. Solve for the eigenvectors associated with the zero eigenvalue:

Take the first row from $H(\theta^*)$ and multiply it by $V$:

$$
\begin{bmatrix}
-\frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} & 0 \\
\frac{2}{\sqrt{2\pi}} \sigma & \frac{n}{\sigma_v^2}
\end{bmatrix}
\begin{bmatrix}
v_k \\
v_k+1 \\
v_k+2
\end{bmatrix}
= \frac{2}{\sqrt{2\pi}} \sigma v_k + \frac{n}{\sigma_v^2} v_{k+2} = 0
$$

$$
\Rightarrow \frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} v_k + \frac{2}{\sqrt{2\pi}} \sigma v_{k+2} = \frac{n}{\sigma_v^2} v_1 + \frac{2}{\sqrt{2\pi}} \sigma v_{k+2} = 0
$$

noting that the first element of $\sum_{i=1}^{n} x_i$ is one $\Rightarrow$

$$
\sum_{i=1}^{n} x = n
$$

$$
\Rightarrow \frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} v_1 + \frac{2}{\sqrt{2\pi}} \sigma v_{k+2} = 0 \Rightarrow v_1 = \hat{\sigma} \frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} v_{k+2}
$$

set $v_{k+2} = 1 \Rightarrow v_1 = \hat{\sigma} \frac{\sum_{i=1}^{n} x_i'}{\sigma_v^2} = \hat{\sigma} \sqrt{\frac{2}{\pi}}$

For this case the intercept ($v_1$) is important as well as $\lambda$.

$$
z_{\gamma} = 
\begin{bmatrix}
\hat{\gamma} \hat{\sigma} \sqrt{\frac{2}{\pi}} \\
0 \\
0 \\
\gamma
\end{bmatrix}
$$

note that $\gamma = \mu$ is the same as Waldman (1982)

**For the N-HN:** $\Delta \ln l = \ln l(\theta^* + z\gamma) - \ln l(\theta^*) = -\frac{\gamma^2 n}{\pi} + \frac{2}{\sqrt{2\pi}} \sigma \ln[2(1 - \Phi(\frac{\gamma^2}{\sigma} - \gamma^2 \sqrt{\frac{2}{\pi}}))]$ (85)

Using a third order Taylor series expansion around the points ($\hat{\beta}$ and $\hat{\lambda} = 0$)
\[ \Delta \ln l \simeq \Delta \ln l(\hat{\beta}, \hat{\lambda} \to 0) + \frac{\partial \ln L}{\partial \beta}(\gamma) + \frac{\partial \ln L}{\partial \lambda}(\gamma) + \frac{1}{2} \left[ \frac{\partial^2 \ln L}{\partial \beta^2}(\gamma)^2 + \frac{\partial^2 \ln L}{\partial \lambda^2}(\gamma)^2 + 2 \frac{\partial^2 \ln L}{\partial \beta \partial \lambda}(\gamma) \right] \]

\[ \frac{1}{6} \frac{\partial^3 \ln L}{\partial \beta^3}(\gamma)^2 + 3 \frac{\partial^3 \ln L}{\partial \beta^2 \partial \lambda}(\gamma)^2 + 3 \frac{\partial^3 \ln L}{\partial \beta \partial \lambda^2}(\gamma)(\gamma)^2 + \frac{\partial^3 \ln L}{\partial \lambda^3}(\gamma)^3 + o(\gamma^4) + o(\gamma^4) \]

All the \( \hat{\beta} \) terms goes to zero when \( \hat{\lambda} = 0 \)

**For the N-HN:** \( \Delta \ln L = \frac{1}{6} \sum c^3 \frac{2}{\sqrt{2\pi}} \left( \frac{\pi - 4}{\pi} \right) \gamma^3 + o(\gamma^4) \) (86)

In conclusion the N-DTN, N-TN and N-TN has stable stationary point at OLS and there is a relationship between the skew of OLS residuals and the MLE estimate of \( \sigma_u \).

## E  Theorem 2

For \( v \sim N(0, \sigma_v^2) \) and \( u \sim \exp(\sigma_u) \). The properties of the Dirac Delta function are exploited in evaluating the N–E model when \( \sigma_u \to 0 \).

### E.0.7 FOC N-E

There are three ways to evaluate the FOCs and SOCs. Firstly, solve the integral and get an analytical solution for the distribution of \( \varepsilon \) and then evaluate FOCs and SOCs when \( \sigma_u \to 0 \). Secondly interchange limit and integral then interchange differentiation and integral and solve the FOCs and SOCs. Thirdly solve the FOCs and SOCs by interchanging differentiation with integral and then interchange limit with the integral. I will proceed with the third way to solve for FOCs and SOCs.

Let \( f_u(u, \theta) \) be a continuous function where \( \theta \) is vector of parameters which includes \( \sigma_u \). Let \( g(u) \) be a continuous function and note that:
1. \[ \lim_{\sigma_u \to 0} \int_0^\infty g(u)f_u(u, \theta)du = \lim_{\sigma_u \to 0} \int_0^\infty \frac{\partial}{\partial \theta} (g(u)f_u(u, \theta))du \text{ (Leibnitz' rule)}. \] Interchange integration with differentiation requires that \( \frac{\partial}{\partial \theta} (g(u)f_u(u, \theta)) \) is integrable.

2. Interchange limit and integral requires that all the sequence of functions and the limit of the function are integrable.

Note from assumption 5 when \( \sigma_u \to 0 \), \( f(u, \sigma_u) \) converges to \( \delta(u) \). The Dirac delta function has mathematical niceties and even though it is not a real function by definition it is integrable.

3. Note that \( E(u) = \sigma_u \Rightarrow E(u) = 0 \) for \( \sigma_u \to 0 \).

The loglikelihood:

\[
\ln l(\sigma_v^2, \sigma_u, \beta; y_i) = \sum_{i=1}^{n} \left[ \ln \left( \frac{1}{\sqrt{2\pi} \sigma_v} \right) - \ln(\sqrt{2\pi}) - \ln(\sigma_v) - \frac{(y_i - x_i'\beta)^2}{2\sigma_v^2} \right] + \ln\left( \int_0^\infty g(v, \beta, u)f_{v\sigma}(u, \sigma_u)du \right) 
\]

Let \( g(\sigma_v, \beta, u) = \exp\left( - \frac{(y_i - x_i'\beta)u}{\sigma_v^2} \right) + \frac{u^2}{2\sigma^2} \frac{1}{\sigma_u} \exp(-\frac{u}{\sigma_u})du \)

The loglikelihood:

\[
\Rightarrow \ln l(\sigma_v^2, \sigma_u, \beta; y_i) = \sum_{i=1}^{n} \left[ -\ln(\sqrt{2\pi}) - \ln(\sigma_v) - \frac{(y_i - x_i'\beta)^2}{2\sigma_v^2} + \ln\left( \int_0^\infty g(\sigma_v, \beta, u)f_{v\sigma}(u, \sigma_u)du \right) \right] 
\]

\[ \text{FOC } \beta: \]

\[
\frac{\partial \ln l}{\partial \beta} = \sum_{i=1}^{n} \left[ \frac{(y_i - x_i'\beta)x_i}{\sigma_v^2} + \int_0^\infty \frac{x_iu g(\sigma_v, \beta, u)f_{v\sigma}(u, \sigma_u)du}{\int_0^\infty g(\sigma_v, \beta, u)f_{v\sigma}(u, \sigma_u)du} \right] 
\]

Let \( g_1(u) = \frac{x_iu}{\sigma_v^2} \Rightarrow g_1(0) = 0 \)
\[
\lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \beta} = \sum_{i=1}^n \left( \frac{y_i - x'_i \beta}{\sigma_v^2} x_i + \frac{\int_0^\infty g_i(u) g(\beta, \sigma_v, u) f_{u_a}(u, \sigma_u) du}{\int_0^\infty g(\beta, \sigma_v, u) f_{u_a}(u, \sigma_u) du} \right)
\]

interchange limit and integral and use the sifting property and the fact that \( g(0) = 1 \) and \( g_i(0) = 0 \)

\[
\lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \beta} = \sum_{i=1}^n \frac{y_i - x'_i \beta}{\sigma_v^2} x_i + \frac{g_i(0) \int_0^\infty \delta(u) du}{g(0) \int_0^\infty \delta(u) du} = 1
\]

\[
\lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_v^2} = \sum_{i=1}^n \left[ -\frac{1}{2 \sigma_v^2} + \frac{y_i - x'_i \beta}{\sigma_v^2} x_i + \frac{\int_0^\infty \left( \frac{(y_i - x'_i \beta) u_i}{\sigma_v^2} + \frac{u_i^2}{2 \sigma_v^4} \right) g(\beta, \sigma_v, u) f_{u_a}(u, \sigma_u) du}{\int_0^\infty g(\sigma_v, \beta, u) f_{u_a}(u, \sigma_u) du} \right]
\]

interchange limit and integral and use the sifting property and the fact that \( g(0) = 1 \) and let \( g_2(u) = \left( \frac{(y_i - x'_i \beta) u_i}{\sigma_v^2} + \frac{u_i^2}{2 \sigma_v^4} \right) \Rightarrow g_2(0) = 0 \) when \( \sigma_u \to 0 \).

\[
\lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_v^2} = \sum_{i=1}^n \left[ -\frac{1}{2 \sigma_v^2} + \frac{(y_i - x'_i \beta)^2}{2 \sigma_v^4} + \frac{\int_0^\infty \left( \frac{(y_i - x'_i \beta) u_i}{\sigma_v^2} + \frac{u_i^2}{2 \sigma_v^4} \right) g(\beta, \sigma_v, u) f_{u_a}(u, \sigma_u) du}{\int_0^\infty g(\sigma_v, \beta, u) f_{u_a}(u, \sigma_u) du} \right]
\]

FOC \( \sigma_v^2 \):

\[
\hat{\beta}_{mle} = (\sum_{i=1}^n x_i x'_i)^{-1} x'_i y = (X'X)^{-1} (X'y) = \hat{\beta}_{ols}
\]

\[
\frac{\partial \ln l}{\partial \sigma_v^2} = \sum_{i=1}^n \left[ -\frac{1}{2 \sigma_v^2} + \frac{(y_i - x'_i \beta)^2}{2 \sigma_v^4} + \frac{\int_0^\infty \left( \frac{(y_i - x'_i \beta) u_i}{\sigma_v^2} + \frac{u_i^2}{2 \sigma_v^4} \right) g(\beta, \sigma_v, u) f_{u_a}(u, \sigma_u) du}{\int_0^\infty g(\sigma_v, \beta, u) f_{u_a}(u, \sigma_u) du} \right]
\]

\[
\lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_v^2} = \sum_{i=1}^n \left[ -\frac{1}{2 \sigma_v^2} + \frac{(y_i - x'_i \beta)^2}{2 \sigma_v^4} + \frac{g_i(0) \int_0^\infty \delta(u) du}{g(0) \int_0^\infty \delta(u) du} = 1 \right]
\]

\[
\frac{\partial \ln l}{\partial \sigma_v^2} = \sum_{i=1}^n \left[ -\frac{1}{2 \sigma_v^2} + \frac{(y_i - x'_i \beta)^2}{2 \sigma_v^4} \right] = \sigma^2_{vMLE} = \sigma^2_{vOLS}
\]

(91)
FOC $\sigma_u$:

$$\frac{\partial \ln l}{\partial \sigma_u} = \sum_{i=1}^{n} \left[ \frac{\int_{0}^{\infty} g(\sigma_v, \beta, u) f_{u,i}(u, \sigma_u) du}{\int_{0}^{\infty} g(\sigma_v, \beta, u) f_{u,i}(u, \sigma_u) du} \right]$$

where $f_{u,i} = \frac{d}{d\sigma_u} f_{u,i} = \frac{1}{\sigma_u} \left( \frac{u}{\sigma_u^2} \right) \exp(-\frac{u}{\sigma_u}) - \frac{1}{\sigma_u^2} \exp(-\frac{u}{\sigma_u}) = \frac{1}{\sigma_u^2} \exp(-\frac{u}{\sigma_u}) [u - \sigma_u]$. Let $f_{u,i1} = \frac{1}{\sigma_u^2} \exp(-\frac{u}{\sigma_u})$ and note that $\lim_{\sigma_u \to 0} [f_{u,i1}] = \frac{1}{\sigma_u} \exp(-\frac{u}{\sigma_u}) = \delta(u)$. Let $g_3(u) = |u - \sigma_u|$ for $u = 0$ and $\sigma_u = 0 \Rightarrow g_3(0) = 0$.

Also recall that $g(0) = 1$

$$\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_u} = \sum_{i=1}^{n} \left[ \lim_{\sigma_u \to 0} \frac{\int_{0}^{\infty} [u - \sigma_u] g(\sigma_v, \beta, u) \left[ \frac{1}{\sigma_u^2} \exp(-\frac{u}{\sigma_u}) \right] du}{\int_{0}^{\infty} g(\sigma_v, \beta, u) f_{u,i}(u, \sigma_u) du} \right]$$

$$\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_u} = \sum_{i=1}^{n} \left[ \lim_{\sigma_u \to 0} \frac{\int_{0}^{\infty} [g_3(u) g(\sigma_v, \beta, u) f_{u,i} du}{\int_{0}^{\infty} g(\sigma_v, \beta, u) f_{u,i}(u, \sigma_u) du} \right]$$

$$\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_u} = \sum_{i=1}^{n} \left[ \int_{0}^{\infty} g_3(0) \int_{0}^{\infty} \delta(u) du \right]$$

$$\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial \ln l}{\partial \sigma_u} = 0$$

(92)

(93)

The first derivative yields OLS at $\theta^* = (\hat{\beta}_{ols}, \hat{\sigma}_{vols}, \hat{\sigma}_{uols} = 0)$.

E.0.8 SOC

For the SOCs the denominator will be written as $g(\sigma_v, \beta, u) f_{u,i}(u, \sigma_u) = \exp(-\left(\frac{(u - \hat{y}'\beta)'u}{\hat{\sigma}_e^2} + \frac{u^2}{\hat{\sigma}_e^2}\right)) f_{u,i}(u, \sigma_u)$.

For $\beta$:
\[
\frac{\partial^2 \ln l}{\partial \beta \partial \beta} = -\frac{\sum_{i=1}^{n} x_i x_i}{\sigma_v^2} + \\
\sum_{i=1}^{n} \left[ \left( \int_0^\infty \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \left( \int_0^\infty \frac{x_i x_i u_i^2}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) - \\
\left( \int_0^\infty \frac{x_i u_i}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \left( \int_0^\infty \frac{x_i u_i}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \right] \\
\left( \int_0^\infty \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right)^2 \right] \\
\] 

(94)

\[
\lim_{\sigma_u \to 0} \left( \frac{\partial^2 \ln l}{\partial \beta \partial \beta} \right) = -\frac{\sum_{i=1}^{n} x_i x_i}{\sigma_v^2} + \\
\sum_{i=1}^{n} \left[ \left( \int_0^\infty \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \left( \int_0^\infty \frac{x_i x_i u_i^2}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) - \\
\left( \int_0^\infty \frac{x_i u_i}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \left( \int_0^\infty \frac{x_i u_i}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right) \right] \\
\left( \int_0^\infty \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2})f_{u_i}(u, \sigma_u)du \right)^2 \right] \\
\] 

(95)

The denominator is 1 and the numerator is zero. For example if we let \( g_1(u) = \frac{x_i x_i u_i^2}{\sigma_v} \exp(-\frac{(y_i-x_i')u_i}{\sigma_v^2} + \frac{u_i^2}{2\sigma_v^2}) \Rightarrow g_1(0) = 0 \) for \( \sigma_u \to 0 \) and note that \( \lim_{\sigma_u \to 0} [f_{u_i}] = \delta(u) \) and recall that \( g(0) = 1 \).

\[
\Rightarrow \lim_{\sigma_u \to 0} \left( \frac{\partial^2 \ln l}{\partial \beta \partial \beta} \right) = -\frac{\sum_{i=1}^{n} x_i x_i}{\sigma_v^2} \\
\] 

(95)

For \( \sigma_v^2 \): Let \( \theta = \sigma_v^2 \)
\[
\frac{\partial^2 \ln l}{\partial \theta^2} = \sum_{i=1}^{n} \left( \frac{1}{2\sigma^2} - \frac{(y_i - x'_i \beta)^2}{\theta^4} \right) + \\
\sum_{i=1}^{n} \left[ \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right]
\]

\[
\left( \int_{0}^{\infty} \left( \left(1 - \frac{2(y_i - x'_i \beta)u_i}{\theta^2} - \frac{u_i^2}{\theta^2} \right) + \frac{u_i^2}{2\theta^2} \right) \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)
\]

\[
\frac{\left( \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)^2}{\sum_{i=1}^{n}} = 1 + \sum_{i=1}^{n} \left( \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)
\]

When we take the limit all the terms equal zero except the first two terms:

\[
\lim_{\sigma_u \to 0} \frac{\partial^2 \ln l}{\partial \theta^2} = \sum_{i=1}^{n} \left( \frac{1}{2\sigma^2} - \frac{(y_i - x'_i \beta)^2}{\theta^4} \right) + \\
\sum_{i=1}^{n} \left[ \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right]
\]

\[
\left( \int_{0}^{\infty} \left( \left(1 - \frac{2(y_i - x'_i \beta)u_i}{\theta^2} - \frac{u_i^2}{\theta^2} \right) + \frac{u_i^2}{2\theta^2} \right) \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)
\]

\[
\frac{\left( \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)^2}{\sum_{i=1}^{n}} = 1 + \sum_{i=1}^{n} \left( \int_{0}^{\infty} \exp\left(-\left(\frac{(y_i - x'_i \beta)u_i}{\theta} + \frac{u_i^2}{2\theta^2}\right)\right) f_{u_n}(u, \sigma_u) \, du \right)
\]

\[
\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial^2 \ln l}{\partial (\sigma_u^2)^2} = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^{n} (y_i - x'_i \beta)^2}{\theta^4} = -\frac{n}{2\sigma_u^2}
\]

For \(\sigma_u\):
\[
\frac{\partial^2 \ln L}{\partial \sigma_n^2} = \Sigma_{i=1}^{\infty} \left[ \int_0^\infty g(\sigma_v, \beta, u)f_{u^2}(u, \sigma_u)du \right] \left[ \int_0^\infty g(\sigma_v, \beta, u)((u - \sigma_u)(\frac{1}{\sigma_n^2} \frac{u^2}{\sigma_u}) \exp(-\frac{u}{\sigma_u}) - \frac{3}{\sigma_u^2} \exp(-\frac{u}{\sigma_u})) \right.
\]
\[
- \frac{1}{\sigma_u^2} \exp(-\frac{u}{\sigma_u})du)
\]
\[
- \int_0^{\infty} g(\sigma_v, \beta, u)(u - \sigma_u)(\frac{1}{\sigma_n^2} \exp(-\frac{u}{\sigma_u}) \int_0^{\infty} g(\sigma_v, \beta, u)(u - \sigma_u)(\frac{1}{\sigma_n^2} \exp(-\frac{u}{\sigma_u})du)
\]
\[
\left( \int_0^{\infty} \exp(-((\frac{u - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right)^2
\]

\[
\Rightarrow \lim_{\sigma_u \to 0} \left[ \frac{\partial^2 \ln L}{\partial \sigma_n^2} \right] = \Sigma_{i=1}^{\infty} \left[ \lim_{\sigma_u \to 0} \left( \int_0^\infty g(\sigma_v, \beta, u)f_{u^2}(u, \sigma_u)du \right) \left( \int_0^\infty \exp(-((\frac{u - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right) \right]
\]
\[
\lim_{\sigma_u \to 0} \left( \int_0^\infty \exp(-((\frac{u - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right)^2
\]

The numerator is zero that is \( g_4(u) = -\sigma_u^2 + (u - \sigma_u)u - 3\sigma_u \Rightarrow g_4(0) = 0 \) when \( \sigma_u \to 0 \) and \( g_3(0) = 0 \) when \( \sigma_u \to 0 \), while the denominator is 1.

\[
\Rightarrow \lim_{\sigma_u \to 0} \left[ \frac{\partial^2 \ln L}{\partial \sigma_n^2} \right] = 0
\]

(99)

For: \( \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_n^2} \)

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_n^2} = \Sigma_{i=1}^{\infty} \left[ \frac{(y_i - x_i^0)(-x_i)}{\sigma_n^2} + \int_0^{\infty} \exp(-((\frac{y_i - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right] \left( \int_0^{\infty} \exp(-((\frac{y_i - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right)
\]
\[
+ \int_0^{\infty} g(\sigma_v, \beta, u)f_{u^2}(u, \sigma_u)du \left( \int_0^{\infty} \exp(-((\frac{y_i - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right)
\]
\[
\left( \int_0^{\infty} \exp(-((\frac{y_i - x_i^0}{\sigma_u})u_i + \frac{u^2}{2\sigma_n^2}))f_{u^2}(u, \sigma_u)du \right)^2
\]

(100)
\[
\lim_{\sigma_u \to 0} \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_v^2} = \sum_{i=1}^{n} \left[ \lim_{\sigma_u \to 0} \left( \int_0^\infty \exp\left( -\frac{(y_i - x_i^\beta)u_i}{\sigma_v} \right) f_{u_x}(u, \sigma_u) \, du \right) \right] - \sum_{i=1}^{n} e_i x_i \sigma_v^4 = 0
\]  

(101)

For Term \( \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_v^2} \):

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_v^2} = \sum_{i=1}^{n} \left[ \int_0^\infty \exp\left( -\frac{(y_i - x_i^\beta)u_i}{\sigma_v} \right) f_{u_x}(u, \sigma_u) \, du \right] \left( \frac{\partial}{\partial \sigma_v} \right) \left( \int_0^\infty \exp\left( -\frac{(y_i - x_i^\beta)u_i}{\sigma_v} \right) f_{u_x}(u, \sigma_u) \, du \right) - \sum_{i=1}^{n} e_i x_i \sigma_v^4
\]

(102)

where \( f'_{u_x} = \frac{d}{d\sigma_v} [f_{u_x}] \).

The numerator is zero and the denominator equals 1 in the limit.
\[
\Rightarrow \lim_{\sigma_u \to 0} \frac{\partial^2 \ln l}{\partial \beta \partial \sigma_u} = 0
\] (103)

For term \( \frac{\partial^2 \ln l}{\partial \sigma_u^2} \)

\[
\left[ \frac{\partial^2 \ln l}{\partial \sigma_u^2} \right] = \sum_{i=1}^{n} \int_0^\infty \exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)du
\]

\[
= \sum_{i=1}^{n} \frac{\exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)}{\left(\int_0^\infty \exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)du\right)^2}
\] (104)

\[
\lim_{\sigma_u \to 0} \frac{\partial^2 \ln l}{\partial \sigma_u^2} = \sum_{i=1}^{n} \lim_{\sigma_u \to 0} \int_0^\infty \exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)du
\]

\[
- \int_0^\infty \exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)du \int_0^\infty \exp\left(-\frac{(y_i - x'_i \beta)u_i}{\sigma_u^2} + \frac{u_i^2}{2\sigma_u^2}\right)f_{u_i}(u, \sigma_u)du
\]

the numerator is zero and the denominator is 1.

\[
\lim_{\sigma_u \to 0} \frac{\partial^2 \ln l}{\partial \sigma_u^2} = 0
\] (105)
F Stable stationary point for the N–E

\[
H(\theta^*) = \begin{bmatrix}
-\frac{\mathbf{x}'\mathbf{x}}{\hat{\sigma}_w^2} & 0 & 0 \\
0 & -\frac{n}{2\hat{\sigma}_w^2} & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

where \( \theta^* = (\hat{\beta}, \hat{\sigma}_w^2, \hat{\sigma}_u = 0) \)

The \( H(\theta^*) \) is negative semi definite with one zero eigenvalue. The eigenvector associated with the zero eigenvalue is:

\[
H(\theta^*) V = 0
\]

\[
\begin{bmatrix}
-\frac{\mathbf{x}'\mathbf{x}}{\hat{\sigma}_w^2} & 0 & 0 \\
0 & -\frac{n}{2\hat{\sigma}_w^2} & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
v_k \\
v_{k+1} \\
v_{k+2}
\end{bmatrix}
= \begin{bmatrix}
0_k \\
0 \\
0
\end{bmatrix}
\]

\[
\Rightarrow \frac{-\frac{\sum x_i x_i'}{\hat{\sigma}_w^2}}{v_k} = 0_k \Rightarrow v_k = 0 \quad (106)
\]

\[
\Rightarrow \frac{-n}{2\hat{\sigma}_w^2} v_{k+1} = 0 \Rightarrow v_{k+1} = 0 
\]

\[
\Rightarrow v_{k+2} 0 = 0 \Rightarrow v_{k+2} = 1 \quad (107)
\]

The eigenvector associate with the zero eigenvalue is:

\[
z = \begin{bmatrix}
0_k \\
0 \\
1
\end{bmatrix} \text{ this imply that } z \gamma = \gamma 
\]

\[
\begin{bmatrix}
0_k \\
0 \\
1
\end{bmatrix} \begin{bmatrix}
0_k \\
0 \\
\gamma
\end{bmatrix} = 0_k
\]

where \( \gamma > 0 \) is a very small positive number. Note that the vector \( z \) has one non-zero element, here the relevant parameter is \( \sigma_u \). The change in the loglikelihood is \( \Delta \ln l = \ln l(\theta^*) - \ln l(\theta^* + z\gamma) \). A third order Taylor series expansion around \( (\hat{\sigma}_u \to 0) \) is used to evaluate if OLS is a stable stationary point in the loglikelihood of the N–E model. If \( l(\theta^* + z\gamma) \geq l(\theta^*) \) then OLS is a stable solution. The change in the loglikelihood evaluated at \( \theta^* \) where we use the fact that \( \sum_{i=1}^n e_i = 0 \).
and $\widehat{\sigma}_v = \frac{\sum_{i=1}^{n} e_i^2}{n}$.

$$\ln l(\theta^n) = \Sigma_{i=1}^{n} \ln \left( \frac{1}{\sqrt{2\pi} \widehat{\sigma}_v} \exp\left( -\frac{1}{2} \frac{(y_i - \widehat{x_i^'} \widehat{\beta})^2}{\widehat{\sigma}_v^2} \right) \right)$$  \hspace{1cm} (109)

$$\ln l(\theta^n + z\gamma) = \Sigma_{i=1}^{n} \ln \left( \frac{1}{\sqrt{2\pi} \widehat{\sigma}_v} \exp\left( -\frac{1}{2} \frac{(y_i - \widehat{x_i^'} \widehat{\beta})^2}{\widehat{\sigma}_v^2} \right) \right) + \Sigma_{i=1}^{n} \ln \left( \exp\left( -\frac{(y_i - \widehat{x_i^'} \widehat{\beta}) u + \frac{u^2}{2\widehat{\sigma}_v^2}}{\gamma} \right) \right) \exp\left( -\frac{u}{\gamma} \right) du$$  \hspace{1cm} (110)

$$\Rightarrow \Delta \ln l = \ln l(\theta^n + z\gamma) - \ln l(\theta^n) = \Sigma_{i=1}^{n} \ln \left( \int_{0}^{\infty} \exp\left( -\frac{(y_i - \widehat{x_i^'} \widehat{\beta}) u + \frac{u^2}{2\widehat{\sigma}_v^2}}{\gamma} \right) \exp\left( -\frac{u}{\gamma} \right) du \right)$$  \hspace{1cm} (111)

The change in the loglikelihood is evaluated using a third order Taylor series expansion around $(\widehat{\sigma}_u \to 0)$.

$$\Delta \ln l \approx \Delta \ln l(\widehat{\sigma}_u) + \frac{\partial}{\partial \sigma_u} \Delta \ln l(\gamma) + \frac{1}{2} \frac{\partial^2}{\partial \sigma_u^2} \Delta \ln l(\gamma) + \frac{1}{6} \frac{\partial^3}{\partial \sigma_u^3} \Delta \ln l(\gamma) + o(\gamma^4)$$

Note that $\ln l(\widehat{\sigma}_u) = 0$, $\frac{\partial \ln l}{\partial \sigma_u} = 0$ (coming from FOC) and $\frac{\partial^2 \ln l}{\partial \sigma_u^2} = 0$ (coming from the SOC and zero eigenvalue/eigenvector condition). All the models discussed have a stable stationary point at OLS with either an identified or unidentified intercept when $\sigma \to 0$. The relationships between the skew of OLS residuals and MLE of $\sigma_u$ for the N-DTN ($\mu \leq 0$), N-TN ($\mu \leq 0$) and N-HN models are due to the fact that the change in the loglikelihoods are a function of $(\sigma_u)$ for all the three models above. Unlike the N-E model the change in the loglikelihood is a function of $(\frac{1}{\sigma_u})$ therefore when $\sigma_u \to 0$, the change in the loglikelihood is zero and hence there is no theoretical relationship between the skew of OLS residuals and the MLE of $\sigma_u$ for this model.

Recall that $g_4(u) = ((u - \sigma_u)u - 3\sigma_u - \sigma_u^2)$ and $g_3(u) = (u - \sigma_u)$ for $\sigma_u \to 0 \Rightarrow u = 0 \Rightarrow g_4(0) = g_3(0) = 0$.
\[
\frac{\partial^3 \ln l}{\partial \sigma_u^3} = \sum_{i=1}^{n} \left[ \frac{2 \left( \int_{0}^{\infty} g(\sigma_v, \beta, u) g_3(u) f_{u\sigma_1}(u, \sigma_u) \, du \left( \int_{0}^{\infty} \frac{1}{\sigma_u} \exp\left(-\frac{u}{\sigma_u}\right) \, du \right) \right) - \left( \int_{0}^{\infty} g(\sigma_v, \beta, u) g_4(u) f_{u\sigma_1}(u, \sigma_u) \, du \right)^2}{\left( \int_{0}^{\infty} \exp\left(-\frac{(y_i - \beta x_i) u}{\sigma_u^2} + \frac{u^2}{2\sigma_u^2}\right) f_{u\sigma}(u, \sigma_u) \, du \right)^2} \right]
\]

The numerator is zero while the denominator is 1.

\[
\lim_{\sigma_u \to 0} \left[ \frac{\partial^3 \ln l}{\partial \sigma_u^3} \right] = 0
\] (112)

\[
\Rightarrow \Delta \ln l = 0
\] (113)

The loglikelihood is stable at OLS for the wrong skew of OLS residuals, however there is no theoretical relationship between the skew of OLS residuals and the MLE of \( \sigma_u \) for the N-E model.

END OF APPENDIX
Essay II: Expected Efficiency Ranks from Parametric Stochastic Frontier Models
Introduction

Given a sample of firm-level data, parametric stochastic frontier models specify production output (or cost) as the sum of a linear response function and an additively composed error, consisting of a two-sided error, representing noise, and a one-sided error, representing inefficiency. See, for example, Aigner, Lovell and Schmidt (1977), Battese and Coelli (1988), Battese and Coelli (1992), and Greene (2005). It is very often assumed that the two-sided error is normally distributed and the one-sided error is truncated normal or exponential. If so, the distribution of inefficiency conditional on the composed error is truncated normal. Given these conditional inefficiency distributions (one for each firm), a common empirical question is how does one assess relative inefficiency in the sample? There are essentially two approaches. The first approach is to calculate the mean of each conditional inefficiency distribution, using the value of the regression residual for each firm in the conditioning argument. See Jondrow et al. (1982) for the cross-sectional case and Battese and Coelli (1988) for the panel data case. These conditional means (evaluated at the residual values) can be ordered across firms, and a sample-wide view of inefficiency is inferred from the order statistic. In particular, the firm with the smallest conditional mean may be deemed efficient relative to the rest in the sample. A second approach is to use the conditional inefficiency distributions to calculate the probability that each firm is best (has lowest inefficiency), conditional on the (joint) composed errors. See Horrace (2005). These conditional efficiency probabilities can be evaluated at the values of the (joint) regression residuals to provide an alternative view of (in)efficiency in the sample, and, in particular, the firm with the largest efficiency probability may be deemed the most efficient. The first approach is a marginal approach in that each conditional mean is derived from a single conditional inefficiency distribution. The second approach is simultaneous in that each
conditional efficiency probability is derived from all the conditional distributions, jointly. In this sense the conditional probabilities contain information from the efficiency rank statistic that the conditional means do not provide. In the parlance of the *multiple comparisons* and *ranking and selection* literatures (e.g., Bechhofer, 1954; Dunnet, 1955; Gupta, 1956, 1955), the conditional efficiency probabilities account for the "multiplicity" in the rank statistic (e.g., firm 1 is better than firm 2 and firm 3 and…).

This paper extends the conditional probability statements of Horrace (2005) to calculate not just the conditional probability that each firm is best (lowest inefficiency), but also the conditional probabilities that each firm is any efficiency rank (best, 2\(^{nd}\) best, …, 2\(^{nd}\) worst, worst) in the sample. The suite of conditional probabilities provide a complete picture of efficiency in the sample and is informative. To see this, let the sample consist of \(n\) firms and let the unconditional distribution of efficiency be the same for each firm (a common assumption). Then, the unconditional probability that any firm is a particular efficiency rank is simply \(1/n\), an uninteresting result. That is, the unconditional probability of any particular efficiency rank can be characterized by a discrete uniform distribution across firms. Once we condition on the sample data (on the regression residuals), the shape of this distribution across firms becomes less uniform (more informative). It is in this sense that the proposed conditional efficiency probabilities are empirically useful. In fact, our simulations show that when the variance of the one-sided error is small relative to that of the two-sided error (a noisy experiment), the conditional probabilities are close to the unconditional result, \(1/n\). As noise decreases, the probability weights of being a particular efficiency rank shift across firms, so the distribution becomes more informative.

Given the suite of conditional efficiency rank probabilities (a partition of the event space
that firm \( i \) is efficiency rank \( r \), it is a simple matter to calculate the expected rank for each firm, conditional on the composed errors, evaluated at the residual values. These conditional expected ranks are also useful. Like the unconditional efficiency rank probabilities, the unconditional expect rank for each firm is constant across firms. For example, if \( n = 5 \) and if the unconditional distribution of inefficiency is (again) identical across firms, then the unconditional expected rank for each firm is \( (1 + \ldots + 5)/5 = 3 \), an uninteresting result. The conditional expected rank, however, varies across firms, and this variability informs our understanding of the efficiency rankings. Continuing the example, if the firm with the highest efficiency score has a conditional expected rank of 1.2 (1 being the best and 5 being the worst), we are much more confident that it is the best firm in the sample than if it has a conditional expected rank of 2.2, and the conditional expected rank of 1.2 is certainly more informative than its unconditional expected rank of 3. Not surprisingly, the informativeness of the conditional expected rank is increasing in the signal to noise ratio in our simulations. Continuing the example, the conditional expected rank of 1.2 for the firm with highest inefficiency score might be from a less noisy experiment than the 2.2 result. In a very noisy experiment the same conditional expected rank might be close to 3, the unconditional result. Our simulations also reveal interesting relationships between the skew of the one-sided error and the distribution of the conditional expected ranks across firms.

This paper is organized as follows. The next section presents the parametric frontier model, the conditional efficiency rank probabilities, and the conditional expected rank measure. The model allows for unbalanced panels, a case which has not been treated extensively in previous work on efficiency probabilities. In section 3, a Monte Carlo study demonstrates how the empirical distribution of conditional efficiency rank probabilities and the conditional expected ranks vary with a) the amount of noise in the experiment and b) the skew of the
unconditional inefficiency distributions. Section 4 presents an empirical application to vessel efficiency in the US North Atlantic Herring fleet, and section 5 concludes.

2. Conditional Inefficiency Rank Probabilities For Parametric Frontiers

We consider the parametric stochastic frontier model for an unbalanced panel of firms:

$$y_{it} = \alpha + x_{it}\beta - u_i + v_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T_t. \quad (1)$$

Here, $y_{it}$ is the observed logarithm of output of the $i^{th}$ firm in the $t^{th}$ period, the $x_{it}$ are observed production inputs, the $u_i \geq 0$ are iid unobserved errors representing unobserved inefficiency, and the $v_{it}$ are iid unobserved errors that cause the efficiency frontier to be stochastic. We assume that the distribution of $v_{it}$ is $N(0, \sigma_v^2)$ and distribution of $u_i$ is the truncation below zero of a $N(\mu, \sigma_u^2)$ random variate.\(^1\) Other distributions for $u_i$ have been considered (e.g., Greene, 1990), but are beyond the scope of what follows. We also require that $x_{it}, u_i$ and $v_{it}$ be independent. Since $y_{it}$ is in log points, firm-level technical efficiency is defined as

$$TE_i = \exp(-u_i).$$

Maximum likelihood estimation of the model's parameters ($\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\sigma}_v^2$ and $\hat{\sigma}_v^2$) is consistent (as $n \to \infty$ or as $T_t \to \infty$).

The model in (1) is fairly flexible. It can represent both Cobb-Douglas and trans-log specifications, and it can be recast as a cost, revenue or profit function. Generalizations for time-varying $u_i$ are plentiful. For example, see Kumbhakar (1990), Cornwell, Schmidt, and Sickles (1990), Battese and Coelli (1992), Lee and Schmidt (1993), Cuesta (2000), Han, Orea and Schmidt (2005), Lee (2005), and Ahn, Lee and Schmidt (2007). Our empirical example in

\(^1\) A fixed effect model is also considered in Schmidt and Sickles (1984).
section 4 involves a more flexible form than in (1), where the marginal products are allowed to vary across groups of firms; the model is estimated using the El-Gamal and Grether (1995, 2000) estimation classification algorithm.

Based on our assumptions in (1), Battese and Coelli (1988) show that the distribution of $u_i$ conditional on the composed error $\varepsilon_i = v_i - u_i$ is the truncation below zero of a $N(\mu_{u_i}, \sigma_{u_i}^2)$ random variate with,

$$\mu_{u_i} = (-\sigma_u^2 T_i \bar{e}_i + \mu \sigma_v^2) / (T_i \sigma_u^2 + \sigma_v^2), \quad \bar{e}_i = T_i^{-1} \sum_{i=1}^{T_i} e_{it} \quad \text{and} \quad \sigma_{v_i}^2 = \sigma_u^2 \sigma_v^2 / (T_i \sigma_u^2 + \sigma_v^2).$$

That is, the conditional density function of $u_i$ is:

$$f(u \mid \varepsilon_i) = \frac{(2\pi \sigma_{u_i}^2)^{-1/2}}{1 - \Phi(-\mu_{u_i} / \sigma_{u_i})} \exp\left\{-\frac{(u - \mu_{u_i})^2}{2\sigma_{u_i}^2}\right\} \text{ with } \varepsilon_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]'.

Then the conditional distribution function is:

$$F(u \mid \varepsilon_i) = \int_0^u f(u \mid \varepsilon_i) \, du = \Phi\left(\frac{|u - \mu_{u_i}| \sigma_i}{\sigma_u} \right) - \Phi\left(-\frac{-\mu_{u_i} \sigma_i}{\sigma_u}\right),$$

where $\Phi$ is the cumulative distribution function of a standard normal random variate. Then the conditional mean of $u_i$ is:

$$E(u \mid \varepsilon_i) = \mu_{u_i} + \sigma_{u_i}^2 \frac{\Phi(-\mu_{u_i} / \sigma_{u_i})}{\Phi(\mu_{u_i} / \sigma_{u_i})},$$

with $\phi$ the density of a standard normal random variate.

In principle, population efficiency ranking is in terms of $u_i$. That is, $u_{[1]} \leq u_{[2]} \leq \ldots \leq u_{[n]}$, so that firm $[1]$ is most efficient in the population, and firm $[n]$ is least efficient. However, $u_i$ is unobserved and cannot be directly estimated, so what is often done is to calculate the vector of residuals $e_i = [e_{i1}, \ldots, e_{iT}]'$ with $e_{it} = y_{it} - \hat{\alpha} - x_{it} \hat{\beta}$, and estimate inefficiency as
\( \hat{u}_i = E(u | \epsilon_i = e_i) \), the conditional mean evaluated at \( \epsilon_i = e_i \) (with \( \hat{\mu} = \mu \), \( \sigma^2_u = \hat{\sigma}_u^2 \) and \( \sigma^2_v = \hat{\sigma}_v^2 \)) for each firm. Empirical exercises often include a rank ordering of the \( \hat{u}_i \), which serves as a predictor of the ordered \( u_i \) conditional on \( \epsilon_i \). If \( T_i = T \), so that \( \sigma^2_i = \sigma^2 \), then the firm rankings based on \( \hat{u}_i \) will be identical to the rankings based on \( \epsilon_i \). However, in the case of unbalanced panels, it is possible that the rankings will not be identical, because \( \sigma^2_i \) causes \( \hat{u}_i \) is no longer be monotonic in \( \epsilon_i \).

Based on the distribution of \( u_i \) conditional on \( \epsilon_i \), Horrace (2005) calculates the conditional probabilities that firm \( i \) is most efficient in the sample, \( \Pr(i = [1] | \epsilon_1, \ldots, \epsilon_n) \), and least efficient in the sample \( \Pr(i = [n] | \epsilon_1, \ldots, \epsilon_n) \). The conditional efficiency probabilities are predicted by evaluating them at \( \epsilon_i = e_i \), \( i = 1, \ldots, n \). We generalize those results to calculate the conditional probability that firm \( i \) is any efficiency rank, \( r \), in the sample, \( \Pr(i = [r] | \epsilon_1, \ldots, \epsilon_n) \). This conditional rank probability is the sum of the probabilities of all possible events where \( u_i \) is rank \( r \) (and there may be many), however it is not necessary to calculate all possible event permutations to determine it. Instead we can start from events in which a particular set of firms are more efficient than \( i \) and the remaining firms are less efficient without regard for the rankings of the firms within each set. Consider one such event with firm \( i \) at rank \( r \), and define subsets \( N_i^- (r) = \{ j : u_j < u_i \} \) and \( N_i^+ (r) = \{ j : u_j > u_i \} \), conditional on \( \epsilon_1, \ldots, \epsilon_n \). The conditional probability of this event is,

\[
\int_0^\infty f(u | \epsilon_i) \prod_{j \in N_i^-(r)} F(u | \epsilon_j) \prod_{k \in N_i^+(r)} [1 - F(u | \epsilon_k)] \, du.
\]

For any rank, except \( r = 1 \) or \( r = n \), there are multiple combinations of ranked firms above and
below $i$ which yield the same rank. In fact, there are
\[ \binom{n-1}{r-1} \] combinations that have
firm $i$ at rank $r$ out of $n$ firms. Accordingly, we index the set of firms above and below $i$ for
each different combination that produces the same rank as $N_i^l(r)$ and $N_i^{l+}(r)$, $l = 1, \ldots, c_{r-1}$.  
Then the conditional efficiency rank probability for rank $r$ out of $n$ is,

\[
\Pr(i = [r] | \varepsilon_1, \ldots, \varepsilon_n) = \sum_{l=1}^{c_{r-1}} \int_{0}^{\infty} f(u | \varepsilon_i) \prod_{j \in N_i^l(r)} F(u | \varepsilon_j) \prod_{k \in N_i^{l+}(r)} [1 - F(u | \varepsilon_k)] \, du , ~ (2)
\]

$i = 1, \ldots, n$, $r = 1, \ldots, n$. When $r = 1$ or $r = n$ these reduce the conditional efficiency probabilities
of Horrace (2005). The $n^2$ probabilities in (2) can be predicted by evaluating them at $\varepsilon_i = \varepsilon_i$, 
$i = 1, \ldots, n$ (with $\hat{\mu} = \mu$, $\sigma_u^2 = \hat{\sigma}_u^2$ and $\sigma_v^2 = \hat{\sigma}_v^2$). It is not difficult to generate computer
algorithms for efficient calculation of these probabilities. When $n$ is large, numerical calculation
of the probabilities may be difficult, but they could certainly be estimated using resampling
techniques.

If we substitute the unconditional density function, $f(u)$, and distribution function, $F(u)$,
for the conditional density function and distributions (respectively) in (2), then it is clear that
$\Pr(i = [r]) = \Pr(j = [r])$ with $\sum_i \Pr(i = [r]) = 1$, so that $\Pr(i = [r]) = 1/n$. This argument hinges
on the unconditional draws of $u$ (over $i$) being identically distributed. Obviously, if the
unconditional distribution of the $u_i$ varies over $i$ (e.g., Battese and Coelli, 1995), then the
unconditional $\Pr(i = [r])$ would not equal $1/n$ in general, and would be a function of the
parameters of the underlying unconditional distributions. Also, if $\sigma_v^2$ is large relative to $\sigma_u^2$, then
realizations of $\varepsilon_i$ contain relatively little information about $u_i$, so that the conditional
distribution of $u_i$ is close to its unconditional distribution. Therefore, when $\sigma_v^2$ is large relative
to $\sigma^2_u$, the probabilities in (2) will be close to $1/n$.

Of course reporting the $n^2$ probabilities in (2) in an empirical exercise may be impractical. However, much of the pertinent information contained in the $r = 1, \ldots, n$ conditional rank probabilities for a firm can be summarized with its conditional expected rank statistic, $\rho_i = \sum_{r=1}^{n} r \Pr(i = [r] | \varepsilon_1, \ldots, \varepsilon_n) \in [1, n]$, (3)
i = 1, \ldots, n$. This measure is an alternative way to characterize efficiency ranks that accounts for multiplicity in the rank statistic through the probabilities in (2). Again, it can be predicted by evaluating $\varepsilon_i$ at the values of $e_i$ for every firm. It also responds to the relative magnitudes of the signal ($\sigma^2_u$) and noise ($\sigma^2_v$) in the same way as the probabilities in (2). In a particularly noisy setting, the conditional rank probabilities in (2) are approximately equal to $1/n$, and the conditional expected rank will be approximately equal across firms. In this sense (2) and (3) provide information on one source of uncertainty in the efficiency ranks that the conditional means, $E(u | \varepsilon_i)$, do not. Of course, the conditional means, the conditional rank probabilities, and the conditional expected rank are all different measures, so comparisons of their abilities to serve as substitutes should not be overstated.

All of the different characterizations of inefficiency (and their relative rankings) are evaluated at $\varepsilon_i = e_i$. Therefore, they all ignore estimation error, which (of course) is asymptotically negligible. Nonetheless, it may be important in finite samples. For the conditional means, there are ways to address the issue. Simar and Wilson (2009) and Wheat, Smith and

\footnote{One could supplement the conditional means with the conditional prediction intervals of Horrace and Schmidt (1996), to judge how much the marginal distributions overlap. The degree of overlap may correspond to the extent to which the conditional probabilities and expected ranks are close to their unconditional counterparts, but this might be highly subjective and (perhaps) lead to an inaccurate assessment of the nature of efficiency in the population.}
Greene (2013) recommend resampling techniques to incorporate estimation error into confidence
intervals on technical inefficiency. Resampling techniques could certainly be employed to assess
the effects of estimation error on the conditional expected ranks. The procedure to do so would
be straightforward, but this is not the focus of the evaluation presented below.

3. Monte Carlo Study
We use a series of simulations to demonstrate properties of the conditional expected rank
statistic. For simplicity, we always set $T_i = 1$. As equation 2 shows, the conditional rank
probabilities for each firm depend on the conditional distributions, $f(u | \varepsilon_i)$, which themselves
depend on three parameters $\mu, \sigma_u^2$ and $\sigma_v^2$. First, we follow standard simulation practice for
stochastic frontier models (e.g., Olson, Schmidt and Waldman, 1980), and explore how statistical
noise, $\sigma_v^2$, affects the empirical distribution of the conditional rank probabilities in (2) and the
conditional expected ranks in (3). To this end we consider $\sigma_v^2 = \{0.01, 0.1, 1, 10\}$ for fixed $V(u)$
. The point is that increasing noise should degrade the efficiency rank probabilities' ability to
accurately detect the true rank of any firm, so that the conditional expected ranks are increasingly
uninformative. Second, Feng and Horrace (2012) show that the skew of the inefficiency
distribution can also confound our detection of firm ranks at different ends of the order statistic
in different ways.\footnote{Feng and Horrace are only concerned with detecting the best firm. We want to detect the rank of all firms.} If the inefficiency distribution is "mostly stars" having many firms in the left
tail ($u_i \equiv 0$ with high probability), then it is difficult to differentiate the individual ranks of these
highly efficient firms (low $[r]$ firms). Conversely, if the inefficiency distribution is "mostly dogs"
having fewer firms in the left tail ($u_i \equiv 0$ with low probability), then it is easier to differentiate
the individual ranks of these highly efficient firms.\textsuperscript{4} The amount of relative mass in one tail of a distribution affects the skew of the distribution. Therefore, our second interest is in seeing the effects of distributional skew (for a fixed variance) on the conditional rank probabilities and the conditional expected ranks. To do this we select values $\mu$ and $\sigma_u^2$ that hold the variance constant at $V(u) = 0.36$ (the variance of a standard normal random variable truncated at zero) but produce skewnesses of 0.5, 1.0, and 1.5 respectively.\textsuperscript{5} These values are listed in Table 1.

The different combinations of parameters ($\mu$, $\sigma_u^2$, $\sigma_v^2$) yield a total of 12 separate exercises (four exercises for each of three skew levels). In each exercise we use a total of 5,000 replications. We use a modest number of firms, $n = 5$, to reduce the computational burden in (2) and simplify exposition.\textsuperscript{6} We ignore the frontier specification and simulate the model:

$$\varepsilon_i = v_i - u_i,$$

so we are implicitly assuming that the production function is known. Our interest is not to understand how well the stochastic frontier model in (1) can be estimated, for this is widely known (e.g., Olson, Schmidt and Waldman, 1980). It is simply to demonstrate the empirical utility of the proposed conditional rank probabilities and the conditional expected rank statistic, and to examine their responses to changes in noise and skew.

Results are shown in Figures 1, 2, and 3 for skew equal 0.5 (low), 1.0 (medium), and 1.5 (high), respectively. We couch our discussion on the effects of changes in $\sigma_v^2$ in terms of Figure 1 (low-skew, $Skew(u) = 0.5$), but it could equally apply to Figures 2 and 3. To achieve $Skew(u) = 0.5$ while holding $V(u) = 0.36$, Table 1 shows that we select $\mu = 0.89$, $\sigma_u^2 = 0.52$ for the results in Figure 1. The figure contains four panels, corresponding to each of four different

\textsuperscript{4} The nomenclature "mostly stars and dogs" is due to Qian and Sickles (2008).

\textsuperscript{5} The skew of a truncated normal is necessarily positive. We use the "standardized" definition of skewness where the 3\textsuperscript{rd} central moment is divided by the third power of the standard deviation.

\textsuperscript{6} Again, the probabilities in (2) could be easily simulated for large $n$, but for the purposes of illustration, small $n$ is sufficient.
values of noise, $\sigma_v^2 = \{0.01, 0.1, 1, 10\}$. Each panel is read similarly. Consider the upper-left panel of Figure 1 where $\sigma_v^2 = 0.01$. For each of 5 firms we have the average of the conditional rank probabilities computed in each replication: probability of rank 1 (dark blue), probability of rank 2 (red), probability of rank 3 (green), probability of rank 4 (purple), and probability of rank 5 (light blue). The population ranks are assigned in each replication such that $i = [r]$. That is, by design firm 1 is $1^{st}$ most efficient in the sample, firm 2 is $2^{nd}$ most efficient in the sample, ..., and firm 5 is least efficient in the sample for any of our 5,000 Monte Carlo draws. These firm numbers, $i$, are along the x-axis, and the average conditional rank probabilities,

$$\Pr(i = [r] | \varepsilon_1, ..., \varepsilon_n),$$

are along the y-axis on the graph. Consider $\Pr(i = [1] | \varepsilon_1, ..., \varepsilon_n)$, the dark blue series. Based on $\sigma_v^2 = 0.01$, the probability that firm 1 is rank 1, $\Pr(1 = [1] | \varepsilon_1, ..., \varepsilon_n)$, is 0.820; the probability that firm 2 is rank 1, $\Pr(2 = [1] | \varepsilon_1, ..., \varepsilon_n)$, is 0.151; the probability that firm 3 is rank 1, $\Pr(3 = [1] | \varepsilon_1, ..., \varepsilon_n)$, is 0.026; and $\Pr(4 = [1] | \varepsilon_1, ..., \varepsilon_n) = \Pr(5 = [1] | \varepsilon_1, ..., \varepsilon_n) = 0$. The probabilities for the other series are read from the graph similarly. For example, for $\sigma_v^2 = 0.01$ we have: $\Pr(2 = [2] | \varepsilon_1, ..., \varepsilon_n) = 0.673$ (red), $\Pr(3 = [3] | \varepsilon_1, ..., \varepsilon_n) = 0.662$ (green), $\Pr(4 = [4] | \varepsilon_1, ..., \varepsilon_n) = 0.716$ (purple), and $\Pr(5 = [5] | \varepsilon_1, ..., \varepsilon_n) = 0.867$ (light blue). For the lowest noise experiment ($\sigma_v^2 = 0.01$), we see that the analysis is better at correctly detecting firms with higher true $[r]$ than firms with lower true $[r]$ (compare $\Pr(1 = [1] | \varepsilon_1, ..., \varepsilon_n) < \Pr(5 = [5] | \varepsilon_1, ..., \varepsilon_n)$ and $\Pr(2 = [2] | \varepsilon_1, ..., \varepsilon_n) < \Pr(4 = [4] | \varepsilon_1, ..., \varepsilon_n)$), and this is always the case for our simulations (regardless of skew), because the distribution of $\nu$ will always have a thinner right tail (where $[r]$ is large) than left tail (where $[r]$ is small) as the skew of a truncated normal is always positive. However, the low skew (0.50) of the Figure 1 simulations means that the
distribution is relatively symmetric, so we shall see that differences in the ability of the conditional rank probabilities to accurately detect high and low ranked firms will become even more stark as we increase the skew (and increase uncertainly over which firms have lower true $r$). See Figures 4, 5 and 6 for a typical empirical inefficiency distribution for each of our three levels of skew: 0.5, 1.0 and 1.5, respectively. Each figure is a kernel density plot using a Gaussian kernel, a Silverman-type bandwidth selection rule, and no boundary-bias correction.

Continuing with the low-skew results of Figure 1, as we increase $\sigma_v^2 = \{0.01, 0.1, 1, 10\}$, the empirical distribution of the efficiency probabilities become more uniform (and less informative). However, we also see in the four panels of Figure 1 that it is always the case that $\Pr(1 = [1] | \varepsilon_1, \ldots, \varepsilon_n) < \Pr(5 = [5] | \varepsilon_1, \ldots, \varepsilon_n)$, even in the noisiest ($\sigma_v^2 = 10$) panel. Both of these empirical phenomena remain as we increase the skew (asymmetry) of the distribution of $u$ to 1.0 and to 1.5 in Figures 2 and 3, respectively (while holding $V(u)$ constant). Looking across the figures we see the effect. Consider the lowest noise panel (upper left panel) in Figures 1, 2 and 3. As the skew increases across Figure 1, 2 and 3, $\Pr(1 = [1] | \varepsilon_1, \ldots, \varepsilon_n)$ is decreasing (0.820, 0.754 and 0.699, respectively), while $\Pr(5 = [5] | \varepsilon_1, \ldots, \varepsilon_n)$ is slightly increasing (0.867, 0.873 and 0.879, respectively). In the words of Qian and Sickles (2008), when the conditional distribution of $u$ has "fewer stars" (low skew of Figure 4) it is easier to detect stars, $\Pr(1 = [1] | \varepsilon_1, \ldots, \varepsilon_n) = 0.820$, than when there are "mostly stars" (high skew of Figure 6), $\Pr(1 = [1] | \varepsilon_1, \ldots, \varepsilon_n) = 0.696$. These are inferential insights that the conditional means, $E(u | \varepsilon_i)$, would not uncover. These are also manifest in the conditional expected ranks which we now consider.

Once the conditional rank probabilities are calculated for each firm at each rank,
calculation of the conditional expected ranks of equation 3 is straight-forward. The distributions
of conditional expected rank (for each simulation run in Figures 1, 2 or 3) are contained in
Tables 2, 3 and 4 (respectively). The utility of the conditional expected ranks is immediately
obvious. First, the extent to which noise affects $\Pr(i = [r] | \varepsilon_1, \ldots, \varepsilon_n)$ is clear. Consider the first
panel ($\sigma_i^2 = 0.01$) of Table 2. The difference between the true rank of firm 1 (first column) and
the average conditional expected rank (second column) is relatively small ($1 - 1.21 = -0.21$), but
this difference is increasing in magnitude as we read down the panels and the level of noise
increases: $1 - 1.71 = -0.72$, $1 - 2.66 = -1.66$, and $1 - 2.96 = -1.96$. These qualitative results are
true for all firms (true rank) and for all levels of skew (Tables 2 thru 4). Obviously, as noise
increases the conditional expected ranks are moving toward the unconditional expected rank, 3
(bottom panel in Table 2), which reflects the nearly uniform distribution of the conditional
efficiency probabilities (bottom panel of Figure 1). Also, the response of the quantiles of the
expected ranks (columns with the heading "Quantiles") to increasing noise is clear: noise tends to
push extreme quantiles (and their surrounding probability mass) to the center of the empirical
distribution of the conditional expected ranks. Second, the effect of skew is clear across the
tables. Consider firms 1 and 5 in the first (low noise) panels of Tables 2 – 4. For firm 1, the
difference between its true rank [1] and the average conditional expected rank is increasing in
magnitude (-0.21, -0.30, -0.39) as skew increases (0.5, 1.0, and 1.5) across Tables 2, 3 and 4,
respectively, while the same differences for firm 5 are non-increasing in magnitude across the
tables (0.16, 0.15, 0.15). Again, this reflects the fact that as skew increases (and there are
relatively more stars in the inefficiency distribution) it is harder to detect "stars" in the left tail of
the inefficiency distribution than to detect "dogs" in the right tail. Third, the conditional
expected ranks are a convenient normalization of relative efficiency. Notice that the
normalization is pegged to both ends of the true order statistic (1 and 5), such that \( \rho_i \in [1, n] \).

Compare this to the traditional predictor of \( \bar{TE}_i = \exp(-u_i) \),

\[
\bar{TE}_i = E[\exp(-u) \mid \varepsilon_i] = \frac{1 - \Phi\left(\frac{\mu_i}{\sigma_i} \right)}{1 - \Phi\left(\frac{-\mu_i}{\sigma_i} \right)} \exp\left\{ -\mu_i + \frac{1}{2} \sigma_i^2 \right\}, \tag{4}
\]

evaluated at \( \varepsilon_i = e_i \). (See Jondrow, Lovell, Materov and Schmidt, 1982.) This absolute predictor normalizes efficiency predictions to the unit interval, \( \bar{TE}_i \in (0, 1) \). Therefore, linear renormalizations of expected rank, like \( 1 - (1 - \rho_i) / n \), can be thought of as alternatives to the \( \bar{TE}_i \) normalization. However, the former is measured on a relative (within sample) scale, while the latter is measured on an absolute (out of sample) scale.

4. Empirical Example

To illustrate our results on expected ranks we revisit the empirical exercise in Flores-Lagunes, Horrace and Schnier (2007), who estimate a stochastic production frontier for an unbalanced panel for \( n = 39 \) vessels from the US North Atlantic Herring fleet (2000-2003). They specify a heterogeneous production function and use the El-Gamal and Grether estimation classification algorithm (El-Gamal and Grether, 1995, 2000) to classify the fleet into three production tiers. See Flores-Lagunes, Horrace and Schnier (2007) for a complete discussion of the data, the production function and the estimation algorithm.\(^7\) Suffice it to say that vessel output is total

\(^7\) The results of the estimation are not reproduced here to focus attention on the different characterization of efficiency ranks and the importance of the proposed conditional expected rank statistic.
catch (tons) and inputs are things like vessel size (tons), hours at sea, and crew size. The estimation yields \( \mu_i \) and \( \sigma^2_{i} \) for each vessel. That is, each vessel's conditional inefficiency distribution is a \( N(\mu_i, \sigma^2_{i}) \) truncated at zero.

The North Atlantic Herring fleet consists of two technologies: trawlers and "purse seiners." While in motion, trawling vessels drag large nets to take catch. A purse seine is a large net that is dropped toward the ocean floor while the vessel is at rest. The gear encircles catch as it is hauled back up to the boat. Vessels use only one of these technology (there are costs to refitting vessels with the different gear types). The El-Gamal and Grether estimation classification algorithm stratifies the fleet into three production tiers, where each tier has separate marginal product estimates (estimates of \( \alpha \) and \( \beta \) in equation 1). The first and second tiers consist exclusively of trawlers and the third tier consists of a mix of trawlers and purse seiners. Efficiency is characterized within (and not across) each production tier.

The estimates of \( \mu_i \) and \( \sigma^2_{i} \) for the five most efficient vessels in each production tier (tier 1, tier 2 and tier 3) are reproduced in the second and third columns of Tables 5, 6, and 7, respectively. The first column contains the unique vessel numbers from the Flores-Lagunes, Horrace and Schnier analysis. The fourth column contains traditional technical efficiency predictors \( \hat{TE}_i = E[\exp(-u) | e_i] \) from (4) evaluated at \( e_i = e_i \), and the results in Tables 5-7 are ranked on this value. The last column contains the conditional expected ranks, \( \hat{\rho}_i \), in (3) for the five most efficient vessels in each tier. The results are compelling. Starting with Table 5, we see that the conditional expected ranks only range in value from 2.272 (vessel 14) to 3.479 (vessel 2), indicating a fairly noisy analysis. Had this been a particularly precise empirical exercise the range would be closer to 1 to 5. One cannot infer this noisiness directly from the \( \hat{TE}_i \), but it is
reassuring to see that the predictor only ranges from 0.846 (vessel 14) to 0.928 (vessel 2), as well. Also the vessel rankings based on $\hat{TE}_i$ match those based on $\rho_i$ in Table 5. However, Table 6 tells a slightly different story. The range of the conditional expected ranks is tighter than in Table 5 and only ranges from 2.535 (vessel 21) to 3.533 (vessel 7), so the analysis is more noisy, however, the ranks based on $\hat{TE}_i$ are different than those based on $\rho_i$. In particular, the ranks of vessel 13 and 12 are reversed, and it is clear why this is the case: the truncated normal distributions (upon which they are based) are vastly different in shape even though the means are approximately the same. That is, $E[\exp(-u) | \varepsilon_{12} = e_{12}] \approx E[\exp(-u) | \varepsilon_{13} = e_{13}] = 0.863$, however the means and variance of the distribution before truncation are extremely different. Compare $\mu_{13} = 0.135$ to $\mu_{12} = -0.518$ and $\sigma^2_{13} = 0.009$ to $\sigma^2_{12} = 0.125$. Vessel 12 has more mass near zero in the distribution of $f(u | \varepsilon_i)$, which is better captured by the expected rank. This underscores the danger of using the conditional means alone to make inferences on ranked technical efficiency scores: they simply do not capture the multiplicity that underlies the ranking.

Table 7 tells an even more nuanced story. The range of the expected ranks are wider: from 1.395 (vessel 3) to 4.151 (vessel 34), so this is the most precise rank statistic of the three, yet there is still some switching in the ranks based on $\rho_i$. In particular the ranks of vessels 33 and 16, and of 30 and 34 are switched. Notice that the differences in $\hat{TE}_i$ for these vessel pairs are not that large, so it is really not surprising that the additional information provided by the conditional rank probabilities might switch the ranking. (This was even more so the case for vessels 13 and 12 in Table 6.) However, it underscores the importance of taking into account multiplicity and noise in any ranking exercise.

Table 7 also includes an additional column with the heading "Trimmed $\rho_i". Sometimes
empiricists will calculate ranked $\tilde{TE}_i$ and, in an ad hoc manner, determine that firms with the highest values are "super-efficient." Super-efficient firms are then dropped from the sample, and the remaining efficiency scores are discussed without re-estimating the production function. Obviously this procedure has no effect on the individual efficiency scores, $\tilde{TE}_i$. It does, however, have implications for the conditional expected rank. In the last column of Table 7, we trim the most efficient vessel (vessel 3) based on its efficiency score, $\tilde{TE}_3 = 0.971$. The rationale is that the distance between its score and the second most efficient vessel, $\tilde{TE}_{33} = 0.934$, is the largest among the most and second most efficient vessels across Tables 5 – 7. In doing so, we have deemed vessel 3 to be "super-efficient." Based on the $\tilde{TE}_i$ scores among the remaining vessels, all the (implied) ranks move up by 1. Vessel 33’s rank moves from 2 to 1, vessel 16’s rank moves from 3 to 2, etc.. The $\tilde{TE}_i$ scores are marginal predictors of efficiency, so all the changes in (implied) ranks are uniform. By contrast, the conditional rank probabilities and, consequently, the conditional expected ranks account for ranking multiplicity and are, therefore, affected in a non-uniform way by dropping a firm (or firms) from the rank statistic. This can be seen in the last two columns of Table 7. The improvement in conditional expected ranks are 0.855, 0.796, 1.000, and 0.954 for vessels 33, 16, 30 and 34, respectively. Consequently, the ordering of the vessels based on $\rho_i$ and "trimmed $\rho_i$" are different. The ordering based on $\rho_i$ is 16, 33, 34 and 30 (best to worst), and the ordering based on "trimmed $\rho_i$" is 33, 16, 34 and 30. This switching of the expected ranks of vessels 33 and 16 underscores the fallacy of trimming "super-efficient" firms without a statistical basis that takes into account noise and the multiplicity implied by the order statistic.
5. Conclusions

We extend the current literature on ranked efficiency scores by defining and proposing the use of conditional efficiency rank probabilities and conditional expected efficiency ranks as a means to provide improved insight into efficiency score rankings. Although our model was fairly restrictive, our results can be more broadly applied than this might indicate. Indeed, there is a broad class of parametric models that yield conditional efficiency distributions that are truncated normal and to which our results directly apply. Additionally, even if the resulting conditional distributions are not truncated normals, our results (and the results of Horrace, 2005) can be adapted to these cases.

We demonstrated nuances of the proposed measures with a Monte Carlo Study. The conditional expected ranks responded in predictable ways to the inherent noisiness of a statistical exercise and to the skewness of the underlying efficiency distribution. While it is generally ignored in empirical applications of the stochastic frontier model, skew is a very important moment to consider in drawing conclusions on ranked efficiency predictors. Our empirical example based on fishing vessels underscores the importance of taking into account multiplicity and noise in any ranking exercise, and the empirical relevance of the conditional rank probabilities and the conditional expected ranks is made clear. We also demonstrated that ad hoc trimming of “super-efficient” firms can lead to incorrect inference on the implied ranks of the remaining firms. It may not be wise to uniformly shift the remaining firms up in the efficiency order statistic.

One potential area of future research is that the OLS residuals are necessarily correlated, so while the conditional inefficiency distribution based on the true regression errors are independent, these distributions based on the residuals are technically not so. It would be...
interesting to see if analytic solutions were forthcoming and the correlation of the residual could be estimated or approximated. It may also be worthwhile to considering resampling techniques to calculate conditional expected rank statistics, so large $n$ will not be problematic, and to estimate confidence intervals for the conditional expected ranks, so that the usual assumption $\hat{\beta} = \beta$ can be relaxed.

It may also be fruitful to explore higher moments of the conditional rank distribution for each firm. We have discussed the conditional expectation of the distribution, but it may be worthwhile to consider the conditional variance of the rank of each firm. Calculating the variance, and any higher moments, would be a straightforward exercise based on the conditional rank probabilities that we have presented. One might speculate that firms with high conditional probabilities of being best and worst would have higher conditional variance of their rank distribution than those with high probability of being in the center of the efficiency rank statistic. The best and worst firms will have more weight in one tail of their conditional rank distributions than firms with higher probability at the median efficiency ranks. However, this remains to be seen.

**References**


Battese, G.E. and T.J. Coelli, 1988, Prediction of firm-level technical efficiencies with a


Gupta, S. S. (1965), On some multiple decision (selection and ranking) rules, Technometrics, 7, 225-245.


Horrace, W.C. and P. Schmidt, 1996, Confidence statements for efficiency estimates from


Wheat, Smith and Green (2013) How confident can we be about confidence intervals for firm specific inefficiency scores from parametric stochastic frontier models. Unpublished, NYU.
Table 1: Simulation Parameters and the Resulting Truncated Moments

<table>
<thead>
<tr>
<th>Underlying $N(\mu, \sigma_u^2)$</th>
<th>Moments of the Truncated Distribution</th>
<th>Range of $(\sigma_u^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($\mu$)</td>
<td>Variance ($\sigma_u^2$)</td>
<td>Mean</td>
</tr>
<tr>
<td>0.89</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>0.00</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>-3.00</td>
<td>2.83</td>
<td>0.67</td>
</tr>
</tbody>
</table>
Table 2. Average and Quantiles of Conditional Expected Rank, Low Skew, Skew(u) = 0.50.

\[
\begin{array}{llllllll}
\sigma_v^2 = 0.01 & & & & & & & \\
\text{True} & \text{Quantiles} & & & & & & \\
\text{Rank} & \text{Average} & 10\% & 25\% & 50\% & 75\% & 90\% & \\
1 & 1.21 & 1.00 & 1.00 & 1.04 & 1.30 & 1.70 & \\
2 & 2.05 & 1.53 & 1.85 & 2.00 & 2.23 & 2.64 & \\
3 & 2.98 & 2.36 & 2.77 & 3.00 & 3.19 & 3.55 & \\
4 & 3.91 & 3.37 & 3.76 & 3.99 & 4.06 & 4.37 & \\
5 & 4.84 & 4.45 & 4.83 & 4.99 & 5.00 & 5.00 & \\
\end{array}
\]

\[
\begin{array}{llllllll}
\sigma_v^2 = 0.1 & & & & & & & \\
\text{True} & \text{Quantiles} & & & & & & \\
\text{Rank} & \text{Average} & 10\% & 25\% & 50\% & 75\% & 90\% & \\
1 & 1.72 & 1.12 & 1.29 & 1.58 & 1.99 & 2.54 & \\
2 & 2.31 & 1.50 & 1.81 & 2.20 & 2.73 & 3.24 & \\
3 & 2.93 & 1.99 & 2.43 & 2.93 & 3.44 & 3.87 & \\
4 & 3.63 & 2.64 & 3.17 & 3.71 & 4.13 & 4.49 & \\
5 & 4.41 & 3.56 & 4.12 & 4.59 & 4.88 & 4.98 & \\
\end{array}
\]

\[
\begin{array}{llllllll}
\sigma_v^2 = 1 & & & & & & & \\
\text{True} & \text{Quantiles} & & & & & & \\
\text{Rank} & \text{Average} & 10\% & 25\% & 50\% & 75\% & 90\% & \\
1 & 2.66 & 2.10 & 2.32 & 2.61 & 2.95 & 3.27 & \\
2 & 2.82 & 2.23 & 2.48 & 2.78 & 3.13 & 3.45 & \\
3 & 2.97 & 2.37 & 2.61 & 2.96 & 3.30 & 3.60 & \\
4 & 3.15 & 2.51 & 2.79 & 3.13 & 3.50 & 3.84 & \\
5 & 3.40 & 2.71 & 3.02 & 3.38 & 3.77 & 4.11 & \\
\end{array}
\]

\[
\begin{array}{llllllll}
\sigma_v^2 = 10 & & & & & & & \\
\text{True} & \text{Quantiles} & & & & & & \\
\text{Rank} & \text{Average} & 10\% & 25\% & 50\% & 75\% & 90\% & \\
1 & 2.96 & 2.75 & 2.84 & 2.95 & 3.07 & 3.17 & \\
2 & 2.98 & 2.77 & 2.87 & 2.98 & 3.09 & 3.19 & \\
3 & 3.00 & 2.79 & 2.88 & 3.00 & 3.11 & 3.21 & \\
4 & 3.02 & 2.80 & 2.90 & 3.02 & 3.13 & 3.24 & \\
5 & 3.04 & 2.82 & 2.93 & 3.04 & 3.15 & 3.26 & \\
\end{array}
\]

Table 3. Average and Quantiles of Conditional Expected Rank, Medium Skew, Skew(u) = 1.0.

<table>
<thead>
<tr>
<th>(\sigma_v^2 = 0.01)</th>
<th>True Rank</th>
<th>Average</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.30</td>
<td>1.00</td>
<td>1.01</td>
<td>1.14</td>
<td>1.48</td>
<td>1.83</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.04</td>
<td>1.44</td>
<td>1.75</td>
<td>2.00</td>
<td>2.26</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.94</td>
<td>2.26</td>
<td>2.70</td>
<td>2.99</td>
<td>3.17</td>
<td>3.58</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.87</td>
<td>3.26</td>
<td>3.73</td>
<td>3.99</td>
<td>4.03</td>
<td>4.34</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.85</td>
<td>4.43</td>
<td>4.85</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\sigma_v^2 = 0.1)</th>
<th>True Rank</th>
<th>Average</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.86</td>
<td>1.20</td>
<td>1.41</td>
<td>1.74</td>
<td>2.19</td>
<td>2.65</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.30</td>
<td>1.53</td>
<td>1.81</td>
<td>2.18</td>
<td>2.72</td>
<td>3.21</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.86</td>
<td>1.94</td>
<td>2.34</td>
<td>2.83</td>
<td>3.34</td>
<td>3.83</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.55</td>
<td>2.50</td>
<td>3.06</td>
<td>3.65</td>
<td>4.07</td>
<td>4.47</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.43</td>
<td>3.48</td>
<td>4.13</td>
<td>4.67</td>
<td>4.93</td>
<td>4.99</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\sigma_v^2 = 1)</th>
<th>True Rank</th>
<th>Average</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.65</td>
<td>2.01</td>
<td>2.25</td>
<td>2.60</td>
<td>2.98</td>
<td>3.34</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.78</td>
<td>2.12</td>
<td>2.39</td>
<td>2.73</td>
<td>3.14</td>
<td>3.52</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.95</td>
<td>2.26</td>
<td>2.53</td>
<td>2.90</td>
<td>3.32</td>
<td>3.70</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.14</td>
<td>2.39</td>
<td>2.70</td>
<td>3.12</td>
<td>3.55</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.48</td>
<td>2.65</td>
<td>3.03</td>
<td>3.49</td>
<td>3.95</td>
<td>4.33</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\sigma_v^2 = 10)</th>
<th>True Rank</th>
<th>Average</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.95</td>
<td>2.67</td>
<td>2.79</td>
<td>2.95</td>
<td>3.10</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.97</td>
<td>2.69</td>
<td>2.82</td>
<td>2.97</td>
<td>3.12</td>
<td>3.26</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.00</td>
<td>2.72</td>
<td>2.84</td>
<td>2.99</td>
<td>3.14</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.02</td>
<td>2.73</td>
<td>2.85</td>
<td>3.01</td>
<td>3.17</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.06</td>
<td>2.76</td>
<td>2.90</td>
<td>3.06</td>
<td>3.22</td>
<td>3.36</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Average and Quantiles of Conditional Expected Rank, High Skew, Skew(u) = 1.5.

| σ^2_v = 0.01 |  |  | Quantiles |  |  |  |  |  |
|--------------|---------------|---------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|
| True Rank    | Average       | 10%           | 25%       | 50%           | 75%           | 90%           | 10%           | 25%           | 50%           | 75%           | 90%           |
| 1            | 1.39          | 1.00          | 1.04      | 1.26          | 1.60          | 1.95          | 1.27          | 1.50          | 1.85          | 2.32          | 2.80          |
| 2            | 2.04          | 1.43          | 1.70      | 2.00          | 2.30          | 2.73          | 1.90          | 2.29          | 2.70          | 3.18          | 3.25          |
| 3            | 2.88          | 2.14          | 2.58      | 2.97          | 3.14          | 3.55          | 2.44          | 2.95          | 3.56          | 4.03          | 4.42          |
| 4            | 3.84          | 3.20          | 3.67      | 3.98          | 4.01          | 4.31          | 3.45          | 4.13          | 4.69          | 4.95          | 5.00          |
| 5            | 4.85          | 4.46          | 4.87      | 5.00          | 5.00          | 5.00          | 4.60          | 4.95          | 5.00          | 5.00          | 5.00          |

| σ^2_v = 0.1 |  |  | Quantiles |  |  |  |  |  |
|--------------|---------------|---------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|
| True Rank    | Average       | 10%           | 25%       | 50%           | 75%           | 90%           | 10%           | 25%           | 50%           | 75%           | 90%           |
| 1            | 1.96          | 1.27          | 1.50      | 1.85          | 2.32          | 2.80          | 2.44          | 2.95          | 3.56          | 4.03          | 4.42          |
| 2            | 2.32          | 1.54          | 1.82      | 2.22          | 2.71          | 3.25          | 2.44          | 2.95          | 3.56          | 4.03          | 4.42          |
| 3            | 2.79          | 1.90          | 2.26      | 2.75          | 3.27          | 3.76          | 2.44          | 2.95          | 3.56          | 4.03          | 4.42          |
| 4            | 3.48          | 2.44          | 2.95      | 3.56          | 4.03          | 4.42          | 3.45          | 4.13          | 4.69          | 4.95          | 5.00          |
| 5            | 4.44          | 3.45          | 4.13      | 4.69          | 4.95          | 5.00          | 4.60          | 4.95          | 5.00          | 5.00          | 5.00          |

| σ^2_v = 1   |  |  | Quantiles |  |  |  |  |  |
|--------------|---------------|---------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|
| True Rank    | Average       | 10%           | 25%       | 50%           | 75%           | 90%           | 10%           | 25%           | 50%           | 75%           | 90%           |
| 1            | 2.63          | 1.90          | 2.17      | 2.55          | 3.00          | 3.48          | 2.51          | 2.69          | 2.91          | 3.15          | 3.39          |
| 2            | 2.76          | 2.00          | 2.29      | 2.70          | 3.18          | 3.62          | 2.52          | 2.72          | 2.94          | 3.19          | 3.42          |
| 3            | 2.90          | 2.12          | 2.41      | 2.85          | 3.33          | 3.82          | 2.56          | 2.73          | 2.97          | 3.21          | 3.44          |
| 4            | 3.14          | 2.29          | 2.63      | 3.11          | 3.62          | 4.06          | 2.60          | 3.03          | 3.58          | 4.13          | 4.55          |
| 5            | 3.57          | 2.60          | 3.03      | 3.58          | 4.13          | 4.55          | 2.64          | 2.85          | 3.09          | 3.34          | 3.58          |

| σ^2_v = 10  |  |  | Quantiles |  |  |  |  |  |
|--------------|---------------|---------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|
| True Rank    | Average       | 10%           | 25%       | 50%           | 75%           | 90%           | 10%           | 25%           | 50%           | 75%           | 90%           |
| 1            | 2.93          | 2.51          | 2.69      | 2.91          | 3.15          | 3.39          | 2.52          | 2.72          | 2.94          | 3.19          | 3.42          |
| 2            | 2.96          | 2.52          | 2.72      | 2.94          | 3.19          | 3.42          | 2.56          | 2.73          | 2.97          | 3.21          | 3.44          |
| 3            | 2.98          | 2.56          | 2.73      | 2.97          | 3.21          | 3.44          | 2.59          | 2.78          | 3.01          | 3.26          | 3.48          |
| 4            | 3.02          | 2.59          | 2.78      | 3.01          | 3.26          | 3.48          | 2.64          | 2.85          | 3.09          | 3.34          | 3.58          |
| 5            | 3.10          | 2.64          | 2.85      | 3.09          | 3.34          | 3.58          | 2.64          | 2.85          | 3.09          | 3.34          | 3.58          |
### Table 5. Heterogeneous Vessel Efficiency Results Tier 1, Sorted on $\overline{TE_i}$

<table>
<thead>
<tr>
<th>Vessel $i$</th>
<th>$\sigma_i^2$</th>
<th>$\mu_i$</th>
<th>$\overline{TE_i}$</th>
<th>$\rho_{i5}$</th>
<th>$\rho_i$</th>
<th>$\rho_i \overline{\epsilon}_{min}$</th>
<th>$\rho_i N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.074</td>
<td>0.928</td>
<td>2.272</td>
<td>3.181</td>
<td>3.3606</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.013</td>
<td>0.03</td>
<td>0.905</td>
<td>2.698</td>
<td>3.829</td>
<td>4.165</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.011</td>
<td>0.094</td>
<td>0.881</td>
<td>3.157</td>
<td>4.576</td>
<td>5.026</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.032</td>
<td>0.04</td>
<td>0.859</td>
<td>3.394</td>
<td>5.031</td>
<td>5.584</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.064</td>
<td>-0.077</td>
<td>0.846</td>
<td>3.479</td>
<td>5.176</td>
<td>5.846</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.345</td>
<td>-1.553</td>
<td>0.844</td>
<td>-</td>
<td>5.092</td>
<td>5.821</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.125</td>
<td>-0.360</td>
<td>0.841</td>
<td>-</td>
<td>5.192</td>
<td>5.962</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>0.039</td>
<td>0.126</td>
<td>0.817</td>
<td>-</td>
<td>5.966</td>
<td>6.829</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.020</td>
<td>0.242</td>
<td>0.781</td>
<td>-</td>
<td>-</td>
<td>7.958</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.183</td>
<td>0.034</td>
<td>0.725</td>
<td>-</td>
<td>6.95</td>
<td>8.498</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.039</td>
<td>0.303</td>
<td>0.731</td>
<td>-</td>
<td>-</td>
<td>8.982</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.026</td>
<td>0.387</td>
<td>0.685</td>
<td>-</td>
<td>-</td>
<td>9.977</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>0.039</td>
<td>1.230</td>
<td>0.298</td>
<td>-</td>
<td>-</td>
<td>12.989</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6. Heterogeneous Vessel Efficiency Results Tier 2, Sorted on $\overline{TE_i}$

<table>
<thead>
<tr>
<th>Vessel $i$</th>
<th>$\sigma_i^2$</th>
<th>$\mu_i$</th>
<th>$\overline{TE_i}$</th>
<th>$\rho_{i5}$</th>
<th>$\rho_i$</th>
<th>$\rho_i \overline{\epsilon}_{min}$</th>
<th>$\rho_i N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>0.011</td>
<td>0.041</td>
<td>0.907</td>
<td>2.535</td>
<td>3.129</td>
<td>3.298</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.039</td>
<td>-0.182</td>
<td>0.903</td>
<td>2.547</td>
<td>3.204</td>
<td>3.200</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.009</td>
<td>0.135</td>
<td>0.864</td>
<td>3.318</td>
<td>4.180</td>
<td>4.345</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.125</td>
<td>-0.518</td>
<td>0.863</td>
<td>3.067</td>
<td>3.954</td>
<td>4.143</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.345</td>
<td>-1.129</td>
<td>0.814</td>
<td>3.533</td>
<td>4.718</td>
<td>5.046</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.095</td>
<td>0.095</td>
<td>0.768</td>
<td>-</td>
<td>5.493</td>
<td>5.952</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.095</td>
<td>0.126</td>
<td>0.758</td>
<td>-</td>
<td>5.669</td>
<td>6.099</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.183</td>
<td>0.332</td>
<td>0.743</td>
<td>-</td>
<td>5.651</td>
<td>6.239</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.055</td>
<td>0.332</td>
<td>0.705</td>
<td>-</td>
<td>-</td>
<td>7.028</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.001</td>
<td>0.692</td>
<td>0.501</td>
<td>-</td>
<td>-</td>
<td>9.716</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.009</td>
<td>0.967</td>
<td>0.382</td>
<td>-</td>
<td>-</td>
<td>11.217</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.002</td>
<td>1.032</td>
<td>0.357</td>
<td>-</td>
<td>-</td>
<td>11.712</td>
<td></td>
</tr>
</tbody>
</table>

### Table 7. Heterogeneous Vessel Efficiency Results Tier 3, Sorted on $\overline{TE_i}$

<table>
<thead>
<tr>
<th>Vessel $i$</th>
<th>$\sigma_i^2$</th>
<th>$\mu_i$</th>
<th>$\overline{TE_i}$</th>
<th>$\rho_{i5}$</th>
<th>Trimmed $\rho_i$</th>
<th>$\rho_i \overline{\epsilon}_{min}$</th>
<th>$\rho_i N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.001</td>
<td>0.001</td>
<td>0.971</td>
<td>1.395</td>
<td>----</td>
<td>1.387</td>
<td>1.442</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>33</td>
<td>0.002</td>
<td>0.065</td>
<td>0.934</td>
<td>2.528</td>
<td>1.673</td>
<td>2.524</td>
<td>2.738</td>
</tr>
<tr>
<td>16</td>
<td>0.008</td>
<td>0.011</td>
<td>0.930</td>
<td>2.500</td>
<td>1.704</td>
<td>2.456</td>
<td>2.753</td>
</tr>
<tr>
<td>30</td>
<td>0.001</td>
<td>0.193</td>
<td>0.825</td>
<td>4.425</td>
<td>3.425</td>
<td>-</td>
<td>5.314</td>
</tr>
<tr>
<td>34</td>
<td>0.023</td>
<td>0.173</td>
<td>0.817</td>
<td>4.151</td>
<td>3.197</td>
<td>3.631</td>
<td>5.189</td>
</tr>
<tr>
<td>35</td>
<td>0.006</td>
<td>0.215</td>
<td>0.808</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.568</td>
</tr>
<tr>
<td>20</td>
<td>0.048</td>
<td>0.179</td>
<td>0.783</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.806</td>
</tr>
<tr>
<td>31</td>
<td>0.001</td>
<td>0.494</td>
<td>0.611</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.783</td>
</tr>
<tr>
<td>29</td>
<td>0.007</td>
<td>0.504</td>
<td>0.606</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.993</td>
</tr>
<tr>
<td>23</td>
<td>0.125</td>
<td>0.604</td>
<td>0.555</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.836</td>
</tr>
<tr>
<td>26</td>
<td>0.006</td>
<td>0.602</td>
<td>0.550</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.137</td>
</tr>
<tr>
<td>22</td>
<td>0.006</td>
<td>0.836</td>
<td>0.435</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.980</td>
</tr>
<tr>
<td>36</td>
<td>0.010</td>
<td>0.913</td>
<td>0.403</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12.511</td>
</tr>
<tr>
<td>25</td>
<td>0.006</td>
<td>1.230</td>
<td>0.293</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.263</td>
</tr>
</tbody>
</table>
Figure 1. Low Skew: Skew(u) = 0.5.

Firm i has sample rank i. Probabilities are for population ranks.
Figure 2. Medium Skew: \( \text{Skew}(u) = 1.0 \).

Firm i has sample rank i. Probabilities are for population ranks.

- \( \sigma_v^2 = 0.01 \)
- \( \sigma_v^2 = 0.10 \)
- \( \sigma_v^2 = 1.0 \)
- \( \sigma_v^2 = 10 \)
Figure 3. Large Skew: Skew(u) = 1.5.

Firm i has sample rank i. Probabilities are for population ranks.

Firm i has sample rank i. Probabilities are for population ranks.
Figure 4. Kernel Density of Truncated Normal Distribution with Low Skew: Skew(u) = 0.5.

Figure 5. Kernel Density of Truncated Normal Distribution with Medium Skew: Skew(u) = 1.0.
Figure 6. Kernel Density of Truncated Normal Distribution with Large Skew: $\text{Skew}(u) = 1.5$. 

$N = 1000$  Bandwidth $= 0.1245$
Essay III: Probability Statements for Stochastic Frontier Models

with Spatial Errors
1 Introduction

Parametric Stochastic Frontier Analysis (SFA) decomposes the error ($\varepsilon = v - u$) into two parts; $v$ accounts for factors uncontrollable by the producers such as bad or good weather while $u$ accounts for inefficiency that is the shortfall in output, see Aigner, Lovell and Schmidt (1977). The composed error ($\varepsilon$) is confounded with noise and signal. Point estimates of inefficiencies are determined post estimation using the conditional mean function which is the first moment of inefficiency conditioned on the composed error term ($E(u|\varepsilon)$), see Jondrow et al.(1982). Firms are then rank based on these point estimates, so that the smallest or largest value of the estimates of $u$ is deemed the most and least efficient firm, respectively, in the sample. Ranking firms based on these estimates of inefficiencies is one of the key importance of SFA, therefore the procedure for determining the most and least efficient firm should not be taken lightly. For a cross section of firms SFA assumes that the random variables, $v_i$ ($i=1,...,n$) and $u_i$ ($i=1,...,n$) are independent as well as the $v_i$ and $u_i$ are independent of each other. This paper relaxes the assumption of independence on $v$ and $u$ but still maintain that they are independent among each other.¹

Horrace and Schmidt (1996) under the assumption that $u_i$ and $v_i$ are independent and identically distributed ($iid$), invert the univariate conditional distributions derived in Jondrow et al. (1982) to develop lower and upper bounds for confidence intervals so that inference could be made about the true value of inefficiency using the predicted values ($\hat{E}(u|\varepsilon) = \hat{u}$). Horrace and Schmidt (1996) also construct confidence intervals using Multiple Comparison with the Best (MCB) technique. This is where lower and upper bounds are developed using Fixed Effect estimates. Accordingly these are joint intervals in which all the firms in the sample are compared to the most efficient or the best firm and there are multiple comparisons simultaneously. This is unlike the Jondrow et al.(1982) confidence intervals which are marginal such that firm $i$

¹For a cross section of firms overtime, Schmidt and Sickles (1984) show how estimates of inefficiencies can be determined without distributional assumption on $v$ and $u$ using a fixed effect estimator. Inefficiencies are assumed to be time invariant and Fixed effect estimators of the slopes are used to compute estimates of the intercepts. Inefficiency for each firm is computed by comparing all the firms in the sample with the most efficient firm that is $\tilde{u}_i = \tilde{\alpha} - \alpha_i$ for ($i = 1,...,n$), where $\tilde{\alpha} = \max(\hat{\alpha}_i)$. Here $\tilde{\alpha}$ is deemed 100% efficient and every other firm is compared to this firm to determined a value for for each $u_i$. However, inference about $\tilde{\alpha}$ is difficult because distributional assumptions on the maximum value yields a non-standard distribution in which the table value for a type I error is difficult to compute.
inefficiency is compared with firm \( j \) inefficiency.\(^2\) One of the drawbacks of the MCB is that these intervals cannot eliminate ties in terms of the maximal and minimal values of inefficiency in the sample. Several firms may have zero as a value for the lower bounds on the confidence intervals and this poses problems in identifying which firm is the best, see Horrace and Schmidt (1996).

Horrace (2005) argues that point estimates (mean of \( (u|\varepsilon) \)) ignores non-trivial information about the characteristics of \( u \) since the mean is not enough to describe the entire distribution of inefficiency, especially when the distribution is asymmetric. Accordingly ranking and selections of firms’ efficiencies should be done based on a probability rule which utilizes both the mean and variance of the distribution of inefficiency and ranks firms simultaneously.\(^3\) Horrace (2005) assumes that the inefficiencies have an independent multivariate truncated normal distribution since firm \( i \) inefficiency is compared with all other firms’ inefficiencies in the sample not necessarily the best firm. The assumption of independence reduces the joint distribution into the products of univariate marginal distributions leading to only a single integral to compute the probability of the least and most efficient firm. These are called probability statements. Because of the \( iid \) assumption on \( u \) firms are indistinguishable so Horrace (2005) substitutes Maximum Likelihood estimates (MLEs) into the univariate conditional distributions \( (u|\varepsilon) \) derived by Jondrow et al. (1982) and then compute the probabilities. The probability statements eliminate ties and hence one is able to say with a certain probability which firm is the least or most efficient.\(^4\) One of the drawbacks of the probability statements is that we are only able to say with a certain probability that firm \( i \) is the least or most efficient, that is we unable to say which firm is the first best, second best or worst in the sample. Horrace, Richards-Shubik and Wright (2015) generalize the probability statements in which conditional rank probabilities are used to calculate expected efficiency ranks as such one is able to say which firm is any efficiency rank from the best, 2nd best, 3rd best,....., 2nd worst or worst in the sample.

\(^2\)For MCB joint confidence intervals are constructed for a vector of differences, such that, \( (u_n - u_1, u_n - u_2, ..., u_n - u_{n-1}) \) where \( u_n \) is the best firm in the sample.

\(^3\)Horrace conditional distribution examines \( (u_i|\varepsilon_1, ..., \varepsilon_n) \) while Jondrow et al. (1982) conditional distribution looks at \( (u_i|\varepsilon) \).

\(^4\)There are enough significant digits in the calculation for the probabilities so ties in ranking should not be problem.
However, the assumption of independence is violated when inefficiencies of neighboring firms are important in the sample. This paper adds spatial correlation to the production function of each firm in a cross-sectional context in which the assumption of independence on \( v \) or \( u \) or both \( v \) and \( u \) are relaxed. This paper makes two assumptions on the inefficiency distribution. First, inefficiency is assumed to be a truncated normal prior to spatial correlation being added to each firm’s production function. Second, the parent distribution of inefficiency is assumed to be normally distributed with spatial correlation and then truncated from the parent distribution. If inefficiency is assumed to be a truncated normal before spatial correlation is added, then the presence of spatial correlation induces an unknown, nonstandard multivariate distribution. This is because a linear combination of truncated normal random variables is not in general a truncated normal, see Horrace (2005). However, if inefficiency is normally distributed prior to truncation then spatial correlation will induce a multivariate truncated normal distribution. Pitt and Lee (1981) model inefficiency for a set of firm using a multivariate truncated normal distribution for the time dimension in a panel data context while this paper models cross-sectional dependence. The dependencies in the \( v \) or \( u \) or both \( v \) and \( u \) result in a likelihood which does not have an analytical form. The likelihood is intractable since the number of integrals increases with the sample size.

This paper generalizes Horrace (2005) probability statements for the maximal and minimal value of \( u \) from \( n \) distributions in the presence of cross-sectional dependence. The presence of spatial correlation induces a variance-covariance matrix with non-zero off diagonal elements and non-constant diagonal elements as such the joint distribution is not the product of the marginal distribution. To compute the probability of the least and most efficient firm in the presence of dependencies requires a multivariate distribution which is onerous since there are multiple integrals involved. This paper uses Sequential Conditioning (SC) to factor the joint distribution into the product of a univariate marginal distribution and several conditional distributions, see Spanos (1986; 1999). According to Spanos (1986; 1999) the key to taming a non-iid sequence of random variables is to apply SC. SC is a technique that is used to reduce the dimensionality of the joint distribution.
For a sample of size \( n \), SC is done by conditioning the \( n^{th} \) random variable on the \( n-1 \) random variables in the sample. For example, if we have three random variables, \( x_1, x_2, \) and \( x_3 \) (the parameter space will be suppressed for now) then the joint distribution is
\[
 f(x_1, x_2, x_3) = f(x_3|x_2, x_1)f(x_2|x_1)f(x_1).
\]
Note that the joint distribution is reduced to univariate conditional distributions and a univariate marginal distribution.

This paper shows that to compute the probability of the least and most efficient firm, SC reduces \( n \) integrals to a single integral. With SC under the assumption of normality the multivariate truncated normal distribution is reduced to a univariate marginal truncated normal distribution and univariate conditional truncated normal distributions which ease the computation of the probabilities of the least and most efficient firm. SC does not help in evaluating the likelihood because it is a function of the unobservables \( u \) and hence not operational. The likelihood has unobservables, the random variable \( u \) and there is no information to determine the values of inefficiencies prior to estimating the likelihood. However, SC helps in computing the probabilities. Post estimation using the conditional mean function, the values of the realizations of the \( u_i \) in the sample can be determined, which are then used for computing the probability statements. Unlike Horrace (2005), if spatial correlation is in the \( u \), MLEs are substituted directly into the univariate marginal distribution and univariate conditional distributions of inefficiency and then probabilities are computed. The conditional distribution of inefficiency conditioned on the composed error \( (u|\varepsilon) \) is not needed to compute the probabilities.

Moreover, if \( u \sim N(0, \sigma_u^2 I_n) \) prior to truncation then SC allows the use of the standard multivariate conditional normal distribution to compute the univariate means and variances that are required to compute the probability statements. SC factors the multivariate truncated normal density into marginal truncated normal and conditional truncated normal densities. The variances are easily picked up from the diagonal elements of the variance-covariance matrix of inefficiency. However, if inefficiencies are drawn from a non-standard, unknown multivariate distribution, (in this case \( n-1 \)) integrals have to be computed to derive the \( n^{th} \) univariate marginal distribution. This becomes computationally involved when the sample size \( n \)
becomes large. If spatial correlation is only in the \( v \) the probability statements reduce to Horrace (2005) in which the conditional distribution of inefficiency conditioned on the composed error \( (u|\varepsilon) \) is used to compute the probabilities. However, cross-sectional dependence will have to be accounted for to make inference more reliable. Additionally this paper demonstrates that the standard conditional mean function that is used to provide an estimate for inefficiency will be incorrect in the presence of spatial correlation since cross-sectional dependence induces a multivariate conditional distribution with heteroskedascity that will distort the ranking of firms’ efficiency.

Over the past decade accounting for spatial correlation or interdependence among firms, states and countries has become an important issue in empirical application, see Lesage and Pace (2009) and Anselin (1988;2001). Kelejian and Prucha (1999) develop a generalized moment estimator to account for the presence of spatial dependence. Lee (2004) derives the asymptotic distribution for Maximum Likelihood and Quasi Maximum Likelihood estimators for the spatial autoregressive model. According to Lee (2004) dependence that exists across spatial units is a relevant issue in urban, real estate, regional, public and industrial organization and to capture spatial dependence, the approaches in spatial econometrics are to impose structures on the model in question. Druska and Horrace (2004) extend the Kelejian and Prucha estimator to the context of SFA where they examine the importance of proximity among several rice farms for a panel data. More recently Glass, Kenjegalieva and Sickle (2014) added the Durbin spatial model to the frontier of a set of firms overtime. They estimate the model using a multivariate search method which uses a concentrated likelihood. Accordingly neighboring firms’ outputs and inputs are important. They also maintain the iid assumption on the error term. Baltagi, Egger and Kesina (forthcoming) examine intersectoral spillovers that affect firms’ productivity in China’s chemical industry in which they used a modified Hausman Taylor approach to capture the spatial correlation in the unobservables.

Spatial correlation is evident at the state, industry and firm level. The decisions to impose a tax rate within a city or a state are influenced by a tax rate charge in a neighboring city or a state. A firm may
decide to locate in an area where there are firms which have similar characteristics, this may lead to better access to suppliers and labor pooling which reduces search cost for finding specialized workers. Additionally firms will locate in places where there are more favorable demand conditions and similar cultural practices, bureaucratic organization, work ethics and economic activities. Having better access to inputs will affect the productivity of a given firm. These activities are not observed by the econometrician, however they affect efficiency and need to be accounted for theoretically and empirically to provide a better characterization of inefficiency. In the presence of spatial correlation or interdependencies it would be unreasonable to assume that the inefficiency distributions are independent. In these contexts the assumption of independence is violated and if not accounted for will lead to inaccurate inference.

The remainder of the paper is divided into six sections following the introduction. Section 2 shows what happen to the composed error when the assumption of independence on \( v \) or \( u \) or both \( v \) and \( u \) are violated. Section 3 adds spatial correlation to the production function of each firm, section 4 discusses identification and the likelihood and provides a brief simulation of the likelihood. Section 5 shows the conditional distribution and section 6 discusses the maximal and minimal draws of inefficiency for \( n \) distributions in the presence of spatial correlation. Section 7 provides a conclusion.

## 2 Violation of the Assumption of Independence on \( \varepsilon \)

SFA models output as a linear function of input, such that, \( y_i = x_i' \beta + \varepsilon_i \), where \( y_i \) is a single output and \( x_i \) is a \( k \times 1 \) vector of inputs with one as the first element. The \( k \times 1 \) vector \( \beta \) is a set of unknown technology parameters to be estimated and \( \varepsilon_i = v_i - u_i \) is the composed error.\(^5\) The \( v_i \in (-\infty, \infty) \) controls for measurement errors and random shocks such as good and bad weather, while the \( u_i \geq 0 \) represents inefficiency for a production function for each firm in the sample.\(^6\) The general assumptions for a cross section of firms are:

---

5The production function may be a Cobb Douglas which is linearized after taking logarithms.

6SFA estimates cost functions, in this case, \( \varepsilon_i = v_i + u_i \).
**Assumption 1:** \( A1. v_i \) for \((i=1,\ldots,n)\) are iid

**Assumption 2:** \( A2. u_i \) for \((i=1,\ldots,n)\) are iid

**Assumption 3:** \( A3. v_i \) and \( u_i \) are independent

**Assumption 4:** \( A4. v_i \) and \( u_i \) are independent of the \( x_i \)

Since the \( v_i \) and \( u_i \) are independent, the composed error \((\varepsilon_i)\) will always be independent because it is the sum of independent random variables, which are independent individually and among each other. If the modeler suspects that there is temporal or spatial dependence on either \( v \) or \( u \) or both \( v \) and \( u \), then this will affect the independence assumptions on the error term, \( \varepsilon \). Before this paper details how to add spatial correlation to the production function of each firm lets examine the validity of the assumption of independence typically imposes on the error term \((\varepsilon)\). Let \( \text{cov}(\varepsilon_i, \varepsilon_k) \) be the covariance between observations \( i \) and \( k \) and impose \( A1-A3 \):

\[
\text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(v_i - u_i, v_k - u_k) = \text{cov}(v_i, v_k) + \text{cov}(v_i, -u_k) + \text{cov}(-u_i, v_k) + \text{cov}(-u_i, -u_k)
\]

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(v_i, -u_k) + \text{cov}(-u_i, v_k) + \text{cov}(-u_i, -u_k)
\]

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(v_i, -u_k) + \text{cov}(-u_i, v_k)
\]

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(v_i, -u_k) + \text{cov}(-u_i, v_k) = 0
\]  

(1)

In arriving at equation 1, the paper first imposes \( A1 \), then \( A2 \) and then \( A3 \), all these three assumptions ensure that the composed error \( \varepsilon_i(i=1,\ldots,n) \) is independent which means that the joint distribution of \( \varepsilon_i \) is equal to the product of the marginal distribution, that is \( f_{\varepsilon_i}(\varepsilon_1,\ldots,\varepsilon_n) = \prod^n_i f_{\varepsilon_i}(\varepsilon_i) \). A point to note is that zero covariance does not imply independence, however independence implies zero covariance. A non-zero covariance implies that the random variables are not independent.\(^7\) Suppose \( u_i \) or \( v_i \) or both \( u_i \) and \( v_i \) are independent means that the joint distribution equals to the product of the marginal distributions.

\(^7\)Independence means that the joint distribution equals to the product of the marginal distributions.
not randomly determined (i.e. there is spatial dependence) then \( A_2 \) or \( A_3 \) or both \( A_2 \) and \( A_3 \) will be too restrictive. How does the violation of \( A_1 \) or \( A_2 \) or both \( A_1 \) and \( A_2 \) affect the covariance structure of the \( \varepsilon_i \)?

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(-u_i, -u_k) \neq 0 \tag{2}
\]

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(v_i, v_k) \neq 0 \tag{3}
\]

\[
\Rightarrow \text{cov}(\varepsilon_i, \varepsilon_k) = \text{cov}(-u_i, -u_k) + \text{cov}(v_i, v_k) \neq 0 \tag{4}
\]

Equation 2 imposes \( A_1 \) and \( A_3 \) while \( A_3 \) is relaxed. Equation 3 imposes \( A_2 \) and \( A_3 \), while \( A_1 \) is relaxed and equation 4 imposes only \( A_3 \). In general we can impose distributional assumptions on the \( u_i \) or \( v_i \) to get an exact expression for the \( \text{cov}(-u_i, -u_k) = -(E(u_i u_k) - E(u_i)E(u_k)) \) or \( \text{cov}(v_i, v_k) = E(v_i v_k) - E(v_i)E(v_k) \).

Given that \( u_i \) or \( v_i \) or both \( u_i \) and \( v_i \) have some dependence structure then the composed error \( (\varepsilon_i) \) which is a function of the \( u_i \) or \( v_i \) will also be affected by this structure. One way of modeling dependencies is by using a multivariate normal density since this captures interdependencies among all the random variables simultaneously, see Kennedy (2003). Note that for equation 4 the covariance \( (\text{cov}(\varepsilon_i, \varepsilon_k)) \) between \( i \) and \( k \) may equal to zero if \( \text{cov}(-u_i, -u_k) = \text{cov}(v_i, v_k) \). Even though there are dependencies in the \( v_i \) and \( u_i \) the dependencies cancel out each other leading to the composed error having a zero covariance structure, but this should not be interpreted as independence and the modeler should still specify a multivariate density for \( \varepsilon \) instead of writing the joint density of \( \varepsilon_i \) as the product of the marginal. This case is beyond the scope of this paper and will not be addressed.

### 3 Frontier Models with Spatial Correlation in the Errors

A standard way of incorporating economic and physical interdependence is by using a prespecified weighted matrix. The most general model for spatial correlation in the errors for a production function is
given in equations 5 to 8. Hereafter the model will be referred to as the Stochastic Frontier Spatial Error Model (SFSEM):

\[
Y = X\beta + \varepsilon \\
\varepsilon = v - u \\
\begin{align*}
    u^* &= \rho W_1 u^* + \delta, \text{ where } \delta \sim_{iid} (0, \sigma_u^2 I_n) \\
v &= \eta W_2 v + \zeta, \text{ where } \zeta \sim_{iid} N(0, \sigma_v^2 I_n)
\end{align*}
\]

This paper uses bold numbers to represent vectors where it is not mention. \(Y\) is a \(n \times 1\) vector of output, \(X\) is a \(n \times k\) matrix of input and \(\varepsilon\) is a \(n \times 1\) vector of unobservables. The prespecified, non-negative \(n \times n\) weighting matrices are \(W_1\) and \(W_2\) and it requires that \(W_1 \neq W_2\) for identification if spatial correlation is modeled in both \(u\) and \(v\) simultaneously. The \(u^*\) and \(v\) are \(n \times 1\) vectors while \(\delta\) and \(\zeta\) are \(n \times 1\) vectors of iid random variables. If \(\delta\) has a half normal distribution then \(u = u^* \geq 0\) while if \(\delta\) is normally distributed then \(u\) is truncated from \(u^* \epsilon (-\infty, \infty)\). Equation 7 assumes that spatial correlation is in the \(u^*_i\) while equation 8 assumes that spatial correlation exits in the \(v_i\). Two additional assumptions are required to ensure that the estimates are consistent:

**Assumption 5 :** A5. The row and column sums of the matrices \(W_1\) and \(W_2\), \((I_n - \rho W_1)\) and \((I_n - \eta W_2)\) before \((W_1\) and \((W_2\) are row-normalized should be uniformly bounded in absolute value as \(n\) goes to infinity.

**Assumption 6 :** A6. The matrices \((I_n - \rho W_1)\) and \((I_n - \eta W_2)\) are non-singular for all the values of \(\rho \in \left(\frac{1}{\rho_{\text{min}}}, 1\right)\) and \(\eta \in \left(\frac{1}{\eta_{\text{min}}}, 1\right)\) where \(\rho_{\text{min}}\) and \(\eta_{\text{min}}\) are the smallest eigenvalue of \(W_1\) and \(W_2\), respectively.

A5 is imposed to ensure that the cross section correlation is limited to a manageable degree, that is the correlation between two spatial units should converge to zero as the distance separating them increases to
infinity, see Keljian and Prucha (1998; 1999). A6 ensures that the \( u_i \) and or the \( v_i \) are uniquely defined. Additionally this paper imposes the following assumption,

**Assumption 7**: \( A7 \). All the diagonal elements of \( W_1 \) and \( W_2 \) are zero.

\( A7 \) means that there is no interaction or dependence between firm \( i \) and itself. The presence of interdependencies induce a multivariate specification. If we suspect that firms are clustering because there are knowledge spillovers or labor pooling (i.e. neighboring characteristics of other firms are important), then we might suspect that the inefficiencies are spatially correlated, see equation 7.

**Result 1** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \delta \overset{iid}{\sim} N(0, \sigma_u^2 I_n) \) the \( n \times 1 \) vector of \( u \) is drawn from a multivariate unknown, nonstandard distribution, see equation 9.\(^8\)

Given equation 7 and the assumption that \( \delta \) is a half normal density, the distribution of \( u \) is:

\[
f(u) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \frac{1}{(\det \Omega_{\sigma_u})^{\frac{1}{2}}} \exp \left( -\frac{1}{2} \left( \mathbf{u} \Omega_{\sigma_u} \mathbf{u} \right) \right)
\]

(9)

where \( \Omega_{\sigma_u} = \sigma_u^2 \left[ (I_n - \rho W_n)' (I_n - \rho W_n) \right]^{-1} \).\(^9\) Result 1 states that if truncation takes place before spatial correlation is added to the model then the vector of \( u \) will have a nonstandard, unknown distribution. This is due to the fact that a linear combination of truncated normal random variables are not necessarily truncated normal distribution, see Horrace (2005). Equation 9 is not a multivariate normal distribution since it has \( \frac{2}{\sqrt{2\pi}} \) instead of \( \frac{1}{\sqrt{2\pi}} \) and neither is it a multivariate truncated normal distribution.\(^{10}\)

**Result 2** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \delta \overset{iid}{\sim} N(0, \sigma_u^2 I_n) \) the \( n \times 1 \) vector of \( u \) is drawn from a multivariate truncated normal distribution, see equation 10.

---

\(^8\)When the distribution is truncated before spatial correlation is added to the model, in order to ensure that \( u \geq 0 \) the weighted matrix \( W_1 \) has to scale such that \( \rho \in [0, 1) \), see Lesage (2009) for a discussion on the different ways of rescaling the weighted matrix.

\(^9\) \( \sigma_u^2 \) is not the variance, it is a scale parameter that is multiply by the variance-covariance matrix of the parent distribution before truncation.

\(^{10}\) The marginal distribution for result 1 is derived in the appendix.
In this case the distribution of $u$ is given by:

$$f(u) = \frac{(\frac{1}{\sqrt{2\pi}})^n \det(\Sigma_{\sigma_u})^{-\frac{1}{2}} \exp(-\frac{1}{2}(u' \Sigma_{\sigma_u}^{-1} u))}{Q_0}$$

(10)

where $Q_0 = \int_0^\infty \cdots \int_0^\infty \left(\frac{1}{\sqrt{2\pi}}\right)^n \det(\Sigma_{\sigma_u})^{-\frac{1}{2}} \exp(-\frac{1}{2}(u' \Sigma_{\sigma_u}^{-1} u)) du_1, \ldots, du_n$

and $\Sigma_{\sigma_u} = \sigma_u^2 [(I_n - \rho W_n)'(I_n - \rho W_n)]^{-1}$. Result 2 states that if truncation takes place after spatial correlation is added to the model then the vector of $u$ will be drawn from a multivariate truncated normal. Note that the marginal distribution of $u_i^* (i = 1, 2) \sim (N(0, \sigma_{u_i}))$ has a non-constant variance given by:

$$f_{u_i^*} = \int_{-\infty}^{\infty} f_{u_{i_1} u_{i_2}}(du_{i_2}) = \frac{1}{\sqrt{2\pi} \sigma_{u_i}} \exp\left(-\frac{u_{i_1}^2}{2\sigma_{u_i}^2}\right)$$

(11)

where $\sigma_{u_i}^2 = \frac{\sigma_u^2(1 + \rho^2 w_{12})}{(1 - \rho^2 w_{12} w_{21})^2}$

and the joint truncated normal distribution:

$$f_{u_1, u_2} = \frac{f_{u_{1_1} u_{1_2}}}{\Pr(u_{1_1} \geq 0, u_{1_2} \geq 0)} = \frac{f_{u_{1_1} u_{1_2}}}{\int_0^\infty \int_0^\infty f(u_{1_1}, u_{1_2}) du_{1_1} du_{1_2}} = \frac{f_{u_{1_1} u_{1_2}}}{Q_0}$$

(12)

where $Q_0 = \int_0^\infty \int_0^\infty f(u_{1_1}, u_{1_2}) du_{1_1} du_{1_2}$. The marginal truncated distribution of $u_{i_1}^*$ is $u_{i_1}$ and it is given by:

$$f_{u_{i_1}} = \frac{1}{Q_0} \frac{1}{\sqrt{2\pi} \sigma_{u_i}} \left(1 - \Phi\left(-\frac{u_{i_1}}{\sigma_u(1 + \rho^2 w_{12})^{\frac{1}{2}}}\right)\right) \exp\left(-\frac{u_{i_1}^2}{2\sigma_{u_i}^2}\right)$$

(13)

The marginal distribution of the bivariate truncated normal distribution is not a truncated normal unless $\rho = 0$, which is similar to the result derived in Arnold et.al (1993) and Horrace (2005). For $i = 1, \ldots, n$ the joint distribution and marginal distributions are more complicated functions of the spatial correlation parameter $\rho$.\(^\text{11}\) Spatial correlation can also be included in the $v_i$, see equation 8. If the rain is expected to fall in

\(^{11}\)If we use SC by Spanos (1986; 1999) the joint distribution is product of a truncated normal and conditional truncated normal distributions, see section 6 and the appendix.
neighborhood A and neighborhood B is close by then given the proximity between the two neighborhoods one will be able to predict if the rain will fall in neighborhood B. This scenario illustrates that the $v_i$ ($i = 1, \ldots, n$) are not randomly determined.

**Result 3** If $\rho = 0$ and $\eta \neq 0$, and $\zeta \sim N(0, \sigma^2 I_n)$ the $n \times 1$ vector of $v$ is drawn from a multivariate normal distribution, see equation 14.

If spatial correlation is in the $v$ then distribution is given by:

$$f(v) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{(\det \Sigma_{\sigma_n})^{\frac{1}{2}}}\right) \exp\left(-\frac{1}{2}(v'\Sigma^{-1}_{\sigma_n}v)\right)$$

(14)

and $\Sigma_{\sigma_n} = \sigma^2_n [(I_n - \eta W_n)'(I_n - \eta W_n)]^{-1}$. If $\rho \neq 0$ and $\eta \neq 0$ then $u$ and $v$ will lead to result 1 and result 3 or result 2 and result 3. In empirical exercise if the modeler is uncertain about which error component (either $v$ or $u$) has spatial correlation then the general model (SFSEM) should be estimated. All the variance-covariance matrices ($\Sigma_{\sigma_n}$, $\Omega_{\sigma_n}$ and $\Sigma_{\sigma_u}$) are symmetric and positive definite matrices. If $\rho = 0$ and $\eta = 0$ then the SFSEM will reduce to the standard model where there is no spatial correlation. From equation 6 and using result 2, A5, A6 and A7 implies that:

$$u^* = [I_n - \rho W_n]^{-1}\delta.$$  

(15)

The distribution of the random vector $u^*$ is a Multivariate Normal distribution of size $n$:

$$u^* \sim N(0, \Sigma_{\sigma_u})$$

(16)

$$E(u^*) = [I_n - \rho W_n]^{-1}E(\delta) = 0$$

$$\Sigma_{\sigma_u} = \sigma^2_n [(I_n - \rho W_n)'(I_n - \rho W_n)]^{-1}$$

(17)
The introduction of spatial correlation induces non-spherical disturbances in which the pretruncated variance-covariance matrix of the parent distribution has non-constant diagonal elements as well as non-zero off diagonal elements. The firms in the sample are affected by geographically proximity and there are variabilities in the characteristics across firms. Let’s examine the SFSEM in scalar notation assuming that $\eta = 0$. For a cross section of firms, equation 19 says that firm $i$ inefficiency is a linear combination of all the neighboring firms’ inefficiencies.

\[
\begin{align*}
y_i &= x_i'\beta + \varepsilon_i & \text{for } (i = 1, \ldots, n) \\
\varepsilon_i &= v_i - u_i \\
u_i^* &= \rho \sum_j w_{ij} u_j^* + \delta_i \\
\delta_i &\sim N(0, \sigma_u^2)
\end{align*}
\]

The $w_{ij}$ are the elements of the prespecified weighted matrix $W$. $\sum_j w_{ij} u_j^*$ is the weighted average of neighboring observations which is used to capture interdependencies among the $n$ observations and $\rho$ measures the strength of the spatial correlation. A positive $\rho$ can be interpreted as economies of scale, while a negative $\rho$ should be interpreted as diseconomies of scale. The agglomeration effects of labor pooling and knowledge spillovers are gains from clustering which means that the production function of each firm will shift outward indicating an expansion in output when $\rho > 0$. If however, there is an overabundant of firms clustering in a specific location then firms will be competing for the factors of production which will increased the demand for inputs, this will increase the cost of clustering relative to the benefit of the positive spillovers and the production function will shift inwards indicating a reduction in output when $\rho < 0$. The interpretation of the marginal effects of the technological parameters remain the same as the standard case since the spatial correlation is in the error term. Let us look at a simple example where there is only one neighboring firm ($i = 1, 2$). Using equation 19 and result 2:
\[ u_1^* = \frac{\rho w_{12} \delta_2 + \delta_1}{1 - \rho^2 w_{12} w_{21}} \]  
\[ u_2^* = \frac{\rho w_{21} \delta_1 + \delta_2}{1 - \rho^2 w_{12} w_{21}} \]  

if there is no spatial correlation then \( u_1^* = \delta_1 \) and \( u_2^* = \delta_2 \) in equation 20 and 21, respectively.\(^{12}\) The variance-covariance matrix is:

\[ \Sigma_{\sigma_u} = \begin{bmatrix} \sigma_u^2 (1 + \rho^2 w_{12}) & \sigma_u^2 \rho (w_{12} + w_{21}) \\ \sigma_u^2 \rho (w_{12} + w_{21}) & \sigma_u^2 (1 + \rho^2 w_{21}) \end{bmatrix} \]

\[ \text{var}(u_1^*) = \frac{\sigma_u^2 (1 + \rho^2 w_{12})}{1 - \rho^2 w_{12} w_{21}} \]  
\[ \text{var}(u_2^*) = \frac{\sigma_u^2 (1 + \rho^2 w_{21})}{1 - \rho^2 w_{12} w_{21}} \]  
\[ \text{cov}(u_1^*, u_2^*) = \frac{\sigma_u^2 \rho (w_{12} + w_{21})}{1 - \rho^2 w_{12} w_{21}} \neq 0 \]  

from the above the covariance between \( u_1^* \) and \( u_2^* \) is not equal to zero unless \( \rho = 0 \) and or \( \sigma_u^2 = 0 \) which implies that the likelihood (\( \varepsilon \)) will be function of a correlation structure induces by spatial correlation which will be discussed in the next section.

### 4 Identification and the Likelihood

For a cross section of firms as \( n \to \infty \) the MLEs of \((\beta, \rho, \eta, \sigma_{\varepsilon}^2, \sigma_u^2)\) are consistently estimated. Since the spatial correlation is in the error term (\( \varepsilon \)), Ordinary Least Squares (OLS) provide consistent estimates for all

\(^{12}\) For \( i = 1, 2 \) and \( \delta_i \overset{iid}{\sim} |N(0, \sigma_u^2)| \) then \( u_1^* = u_1 \) and \( u_2^* = u_2 \) are not guarantee to be non-negative unless \( \rho \in [0, 1) \).
the parameters of the model except the intercept. The intercept is confounded with the spatial correlation parameter and the scale parameter ($\sigma_u^2$) of inefficiency. Corrected OLS (COLS) which is OLS corrects for the bias intercept consistently estimate all the parameters of the model, see Olson, Schmidt and Waldman (1980).\(^{13}\) The distribution of the composed error term ($\varepsilon$) is asymmetric hence MLE is more efficient relative to COLS, furthermore distributional assumptions of the error terms ($v$ and $u$) are required to compute the moments as such the model is typically estimated using Maximum Likelihood technique. If spatial correlation is in the idiosyncratic or noise component ($v$) then OLS consistently estimate all the parameters of the model. In empirical application for a cross sections of firms, the first step is to use COLS or OLS residuals to conduct a Moran I or Lagrange Multiplier test, to test the significant of $\rho$ or $\eta$, respectively, before proceeding to MLEs.\(^{14}\) If we reject the null that $\rho = 0$ or $\eta = 0$, we should specify inefficiency using a multivariate truncated normal distribution or noise using a multivariate normal distribution otherwise we are back in the standard framework where inefficiency or noise is assumed to be iid and follows a univariate truncated normal or a univariate normal distribution, respectively. For numerical optimization of the likelihood, OLS estimates are used as starting values for all the parameters of the model except $\rho$ (zero is used for the starting value for this parameter).

Using result 2 ($\rho \neq 0$ and $\eta = 0$) the covariance between $u_1^*$ and $u_2^*$ is not equal to zero unless any of the three cases occur below. If case 1 occurs we are back in the standard framework where there is no spatial correlation.

$$\text{cov}(u_1^*, u_2^*) = 0 \text{ if } \begin{cases} \text{case 1. } \rho = 0 \text{ and } \sigma_u^2 \neq 0 \\ \text{case 2. } \rho \neq 0 \text{ and } \sigma_u^2 \to 0 \\ \text{case 3. } \rho = 0 \text{ and } \sigma_u^2 \to 0 \end{cases}$$

For case 2, SFA suffers from the "wrong skewness problem"\(^{15}\) which will result into the scale parameter

\(^{13}\)The error term ($\varepsilon$) can transform to have zero mean. Let $\mu_u = E(u)$ and $\alpha^* = \alpha - \mu_u$. $\varepsilon^* = \varepsilon + \mu_u \Rightarrow E(\varepsilon^*) = 0$ and $y = \alpha^* + X\beta + \varepsilon^*$.

\(^{14}\)If spatial correlation is in both $v$ and $u$ then a joint tests will be conducted for $\eta = 0$ and $\rho = 0$, respectively.

\(^{15}\)The total error $\varepsilon = v - u$ and the skew of $\varepsilon$ occurs through $u$ given the standard assumption on $v$ and $u$. The skew of
of the pretruncated variance-covariance matrix of the inefficiency distribution equals to zero, see Olson, Schmidt and Waldman (1980), Waldman (1982), Almanidis Qian and Sickles (2014). Wright (2015) shows that if the wrong skew occurs the distribution of \( u \) converges to a Dirac delta function and the parameters of the inefficiency distribution are not identified. If the wrong skew of OLS residuals occurs then the MLE of \( \sigma_u^2 \) is zero and the variance-covariance matrix of the parent distribution of inefficiency becomes singular and \( \rho \) is not identified. This is very problematic because spatial correlation is present, but Maximum Likelihood technique cannot provide an estimate of \( \rho \). For case 3 there is no spatial correlation and we are back in the standard model. Waldman (1982) provides a solution for the normal-half normal specification when this occurs while Wright (2015) provides a theoretical solution for the normal-truncated normal and the normal-doubly truncated normal distributions when the pretruncated mean is non-positive. This paper will focus on the MLE of \( \sigma_u^2 > 0 \) which corresponds to the correct skew of OLS residuals.

4.1 Likelihood

The likelihood for the composed error with interdependencies among \( n \) firms given the assumptions on \( v \) or \( u \) or both \( v \) and \( u \).

**Result 4** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \delta \overset{iid}{\sim} |N(0, \sigma_u^2 I_n)| \), for a \( n \times 1 \) vector of \( v \) and \( u \), \( f(\varepsilon) \) has a non-standard intractable distribution.

Under \( A1 \) and \( A3 \), the likelihood is:

\( u \) is positive which implies that skew of \( \varepsilon \) is negative for a production function. However, in empirical application because of sampling errors OLS residuals may have a positive skew and this poses problem for MLEs of the inefficiency parameters, this is known as the "wrong skewness problem".

\(^{16}\)All the derivation for result 4 to 6 are in the appendix.
\[ f(\varepsilon) = Z_3(\sqrt{2\pi}) \left( \frac{1}{\sigma_v} \right)^n \left( \frac{1}{\sigma_u} \right)^n \left( \frac{1}{\det(\Omega_{\sigma_v, \sigma_u})} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \left( \frac{\varepsilon}{\sigma_v^2} + \frac{\varepsilon^2}{\sigma_u^2} \right) \right) \]  

\[ Z_3(\varepsilon) = \Pr(u_1 \geq 0, \ldots, u_n \geq 0) = \int_{\varepsilon}^{\infty} \cdots \int_{\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{u + \Omega_{\sigma_v, \sigma_u} \varepsilon}{\sigma_u^2} \right)} \, du_1 \cdots du_n \]

The distribution of \( \varepsilon \) does not have a closed form, it is intractable because the number of integrals \( Z_3(\varepsilon) \) increases with the sample size. This becomes computationally involve when the sample size gets large.

**Result 5** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \varepsilon \overset{iid}{\sim} N(0, \sigma_u^2 I_n) \), for a \( n \times 1 \) vector of \( v \) and \( u \), \( f(\varepsilon) \) has a non-standard intractable distribution.

The likelihood is:

\[ f_\varepsilon(\varepsilon) = \frac{Z_0}{Q_0} \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_v} \right)^n \left( \frac{1}{\sigma_u} \right)^n \left( \frac{1}{\det(\Sigma_{\sigma_v, \sigma_u})} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \left( \frac{\Sigma^{-1}_{\sigma_v, \sigma_u} (u + \Sigma_{\sigma_v, \sigma_u} \varepsilon)}{\sigma_u^2} \right) \right) \]

\[ Z_0 = \int_{0}^{\infty} \cdots \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{u + \Sigma_{\sigma_v, \sigma_u} \varepsilon}{\sigma_u^2} \right)} \, du_1 \cdots du_n \]

\[ Q_0 = \int_{0}^{\infty} \cdots \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{u + \Sigma_{\sigma_v, \sigma_u} \varepsilon}{\sigma_u^2} \right)} \, du_1 \cdots du_n \]

\[ \Sigma_{\sigma_v, \sigma_u} = \left( \frac{I_n}{\sigma_u^2} + \Sigma^{-1}_{\sigma_v} \right)^{-1} \]

The loglikelihood that is used in the simulation is:

\[ \ln l(\rho, \sigma_u^2, \sigma_v^2, y) = \ln(Z_0) - \ln(Q_0) - n \ln(\sqrt{2\pi}) - \ln(\sigma_v) + 0.5(\ln(\det(\Sigma_{\sigma_v, \sigma_u} (\rho, \sigma_u^2))) - 0.5(\ln(\det(\Sigma_{\sigma_v} (\rho, \sigma_u^2)))) + \frac{1}{2\sigma_v^2} (y - X\beta)' \left( I_n + \Sigma_{\sigma_v, \sigma_u} (\rho, \sigma_u^2) \right) \frac{1}{2\sigma_v^2} (y - X\beta) \]
Note that the variance-covariance matrix \((\Sigma_{\sigma,v}\sigma_u)\) is a function of the spatial parameter and the scale parameter of inefficiency and \(Z_0 = Z_0(y - X\beta)\). The violation of \(A1\) or \(A2\) or both \(A1\) and \(A2\) leads to a complicated likelihood function. SC can be used to reduce the dimensionality of the likelihood but it does not help since the marginal distribution and conditional distributions of \(Z_0\) and \(Q_0\) will have unknown realizations of the unobservables \((u)\) which cannot be determined prior to estimating the likelihood.

**Result 6** If \(\rho = 0\) and \(\eta \neq 0\), and \(\zeta \sim N(0, \sigma_v^2 I_n)\), for a \(n \times 1\) vector of \(v\) and \(u\), \(f(\varepsilon)\) has a non-standard intractable distribution.

The likelihood is:

\[
f(\varepsilon) = A_0 \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_u} \right)^n \left( \det \Sigma_{\sigma,v}\sigma_u \right)^{\frac{1}{2}} \exp \left( \frac{1}{2} (\varepsilon^T \Sigma_{\sigma,v}\sigma_u^{-1} \Sigma_{\sigma,v}\sigma_u \varepsilon) \right) \]

\[
A_0 = \text{Pr}(u_1 \geq 0, \ldots, u_n \geq 0) = \int_0^\infty \cdots \int_0^\infty \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \det \Sigma_{\sigma,v}\sigma_u \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \left( u^T \Sigma_{\sigma,v}\sigma_u^{-1} \Sigma_{\sigma,v}\sigma_u u \right) \right) \]

Note that \(A_0\) has integrals which increases with the sample size. If spatial correlation is in both \(v\) and \(u\) (where \(u\) is half normal or \(u\) is normally distributed) jointly then \(f(\varepsilon)\) has a non-standard intractable distribution, see the appendix. This section concludes by stating that the presence of spatial correlation either in the \(v\) or \(u\) or both \(v\) and \(u\) induce an intractable likelihood.

### 4.2 Simulations

This paper uses equation 27 to simulate. A signal to noise were chosen to be, \(\lambda = \frac{\sigma_u}{\sigma_v} = [0.2, 0.5, 1, 2]\) and these correspond to \(\sigma_u = [0.1964, 0.4472, 0.7071, 0.8944]\) and \(\sigma_v = [0.9805, 0.8944, 0.7071, 0.4472]\). The spatial correlation \(\rho = [-0.1, 0.2, 0.4, 0.6]\), a sample size \(n = 3\) (three integrals) and a constant \(\beta_0 = 2\).
This paper sets $\sigma_u^2 + \sigma_v^2 = 1$ and examines the MLE estimates. Inefficiencies are drawn from a multivariate truncated normal distribution in matlab 7.4 version. This program is written with reference to C.P. Robert, an article title, "Simulation of truncated normal variables" published in Statistics and Computing, pp. 121-125 (1995). The model is estimated using Maximum Likelihood Technique. This paper uses a cholesky decomposition to factor the variance-covariance matrix into lower triangular matrices to ease computation numerically. The values reported along the column of Table 1.1 are the mean estimates of $[\beta_0, \sigma_v, \sigma_u, \rho]$. The paper first fixes the spatial correlation ($\rho$) and then allow $\sigma_u$ to vary. It then fixes $\rho$ and allow $\sigma_u$ to vary, see Table 1.1. As we move along the rows of Table 1.1, the parameter $\rho$ is fixed while $\sigma_u$ varies, and along the column $\rho$ varies for fix values of $\sigma_u$. For $\sigma_u = 0.1959$ and $\rho = -0.1$, the mean estimate of $\hat{\rho}$ and $\hat{\sigma}_u$ are $-0.4150$, and $0.3926$, respectively. An increase in $\sigma_u$ of $0.8$ for fix $\rho = -0.1$, the mean estimate of $\rho$ and $\sigma_u$ are $-0.5286$ and $0.1341$, respectively. The estimate for the intercept improves. Simulations confirm that when OLS residuals have the wrong skew the parameters for the inefficiency distribution are not identified. For a value of $\rho = 0.2$ this estimate appears to improve as we increases the signal to noise ratio, a value of $\sigma_u = 0.8$ corresponds to an estimate of $\hat{\rho} = 0.33$.\footnote{In Matlab 7.4 the sample size of a multivariate cumulative distribution function (cdf) is constraint to twenty five (25). As $n$ increases the integrals do not converge and the model does not provide an estimate for $\rho$.}

5 The Conditional Distribution

In SFA post estimation firms are ranked based on point estimates of inefficiencies. Given $A_1$ to $A_4$, the value of inefficiency for each firm $i$ is computed using the conditional mean function that is, $u_i = E(u_i|\varepsilon = \varepsilon_i)$. The univariate truncated normal conditional distribution by Jondrow et al. (1982) is given by:
\[
f(u|\varepsilon) = \frac{1}{\sqrt{2\pi}} \frac{1}{|1 - \Phi_i(\frac{\mu}{\sigma^2})|} \exp(-\frac{1}{2\sigma^2}(u_i + \mu_i)^2) \tag{29}
\]

\[
\mu_i = \frac{\sigma_u^2 \varepsilon}{\sigma_v^2}, \sigma_u = \frac{\sigma_u^2 \sigma_v^2}{\sigma_v^2} \quad \text{and} \quad \sigma^2 = \sigma_u^2 + \sigma_v^2
\]

However in empirical application the MLEs \((\beta, \sigma^2, \lambda = \frac{\sigma^2}{\sigma_v^2})\) and the residuals \(e_i\) are substituted into the conditional mean function to provide an estimate for the true value of inefficiency for each firm \(i\) - \(\hat{u}_i = E(u|\varepsilon = e_i)\).

The presence of spatial correlation induces a multivariate distribution for the vector of \((u|\varepsilon)\). The conditional distribution using results 1 and 4 is:

\[
f(u|\varepsilon) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{Z_3} \frac{1}{(\det \Sigma_{\sigma_v \sigma_u})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(u + \mu_1)^t \Sigma_{\sigma_v \sigma_u}^{-1}(u + \mu_1)\right) \tag{30}
\]

\[
\mu_1 = \Sigma_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}
\]

the conditional distribution using results 2 and 5 is:

\[
f(u|\varepsilon) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{Z_3} \frac{1}{(\det \Sigma_{\sigma_v \sigma_u})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(u + \mu_2)^t \Sigma_{\sigma_v \sigma_u}^{-1}(u + \mu_2)\right) \tag{31}
\]

\[
\mu_2 = \Sigma_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}
\]

the conditional distribution using results 3 and 6 is:

\[
f(u|\varepsilon) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{(\det \Sigma_{\sigma_v \sigma_u})^{\frac{1}{2}}} \frac{1}{A_0} \exp\left(-\frac{1}{2}(u + \mu_3)^t \Sigma_{\sigma_v \sigma_u}^{-1}(u + \mu_3)\right) \tag{32}
\]

\[
\mu_3 = \Sigma_{\sigma_v \sigma_u} \Sigma_{\sigma_v \varepsilon}
\]

---

\(^{18}\)See Horrace (2005) for a derivation of the Characteristic function for a Multivariate truncated normal distribution, from this we can derive the first moment of \((E(u|\varepsilon))\).
Spatial correlation in either \( v \) or \( u \) or both \( v \) and \( u \) result in the conditional distribution \((u|\varepsilon)\) being a multivariate truncated normal distribution. This also true irrespective of the assumption on the inefficiency distribution discusses in the previous sections. If empiricists incorrectly use the univariate conditional distribution derives in Jondrow et al. (1982) to compute point estimates of inefficiencies then this will produce bias values for each \( u_i \). Horrace and Schmidt (1996) invert the conditional distributions by Jondrow et al. (1982) to construct lower and upper bounds for confidence interval for inefficiency. In the presence of spatial correlation the lower and upper bounds have to be modified to account for induce heteroscedasticity and the non-zero off diagonal elements in the variance-covariance matrix. Any inference drawn assuming that there is no spatial correlation will be inaccurate and unreliable unless \( \rho \) and or \( \eta \) is zero.

6 Probability of Maximal and Minimal draws of Inefficiency in the presence of Spatial Correlation

According to Horrace (2005) for iid random variables, comparing inefficiencies between firm \( i \) and \( j \) translates into the probability of the difference between \( u_i \) and \( u_j \) which are not truncated normal random variables and hence any inference drawn assuming otherwise will be inaccurate. The presence of spatial correlation violates the assumption of independence and hence equations 4 and 5 in Horrace (2005) would not be appropriate to compute the probability of the least and most efficient firm. Additionally since the \( u_i \) are assumed iid, Horrace (2005) probability statements have to use the conditional distribution of inefficiency conditioned on the composed error term \((u|\varepsilon_i)\) to compute the probability of the least and most efficient firm. This is because the conditional distribution \((u|\varepsilon_i)\) informs us about variation in the firms’ characteristics for a given population of firms.\(^{19}\)

The presence of spatial correlation implies that the random variables \((v \text{ or } u \text{ or both } u \text{ and } v)\) are

\(^{19}\)Horrace (2005) imposes independence on the \( u_i \) as such: \( \Pr(u_j \leq u_k, j \neq k) = \Pr(u_1 \leq u_k), \Pr(u_2 \leq u_k), ..., \Pr(u_n \leq u_k), \) the multiple integrals are reduce to a single integral.
no longer independent. Since inefficiency is non-random the joint distribution is not the product of the marginal distributions. To use the joint distribution to compute the probabilities of the least and most efficient is an onerous task since the number of integrals increases with the sample size. The method of SC, is one way of reducing the joint distribution of a non-random sample into the product of a marginal and conditional distributions, see Spanos (1986;1999). SC factors the multivariate density into a univariate marginal and the product of conditional densities which are relatively easier in computing the probability of the least and most efficient firm. Spatial correlation induces heteroskedasticity and non-zero off diagonal elements in the variance-covariance matrix which creates variation across each firm in the sample as such we can substitute MLE estimates directly into the univariate marginal and conditional distributions and compute the probabilities. SC reduces the dimensionality curse of computing multiple integrals to a single integral. However, the mean of the univariate conditional distributions is a function of the unobservable \((u_i, i = 1, \ldots, n)\), hence it is not operational. To make the conditional distributions operational the modeler has to first estimate the conditional mean function, \((u = E(u_i|\varepsilon = \hat{\varepsilon}))\) and then substitute these estimates into the mean of the univariate conditional distribution \((u_j|u_{j-1})\), for \(j = 2, \ldots, n\) firms.

Let \(f_{u_k}\) be the probability density function (pdf) of firm \(k\) from the population of firm and let \(F_{u_k}\) be the corresponding cdf. Also let \(f_{u_i|u_{i-1}}\) and \(F_{u_i|u_{i-1}}\) be the conditional density and cdf, respectively, of \(u_i\) conditioned on \(u_{i-1}\). The pdfs and cdfs are absolutely continuous and the standard properties of the cdf holds, \(F(\infty) = 1\) and \(F(-\infty) = 0\). The probability that firm \(k\) is the least and most efficient in the population are given in results 7 to 9 below under different assumptions. The \(k^{th}\) firm is the control index that is firm \(k\) inefficiency is compared to the \(n-1\) firms’ inefficiencies in the population.

**Result 7** If \(\rho = 0\) and \(\eta \neq 0\), and \(\varepsilon \sim iid N(0, \sigma^2 I_n)\), the probability that firm \(k\) is the least and most efficient firm in the population reduces to the Horrace (2005) equation 4 (equation 33 below) and equation 5 (equation 34 below), respectively. The probabilities are computed using the conditional distribution \(f(u|\varepsilon)\) in equation 32.
\[ P(u_j \leq u_k, \ j \neq k) = \int_0^\infty f_k(u_k, \mu_k, \sigma^2_{uk}) \prod_{j \neq k}^{n} F_{u_j}(u_k) du_k \]  
(33)

\[ P(u_j > u_k, \ j \neq k) = \int_0^\infty f_k(u_k, \mu_k, \sigma^2_{uk}) \prod_{j \neq k}^{n} (1 - F_{u_j}(u_k)) du_k \]  
(34)

Given the assumptions on \( v \) and \( u \) (half normal distribution) the pdfs and cdfs in equations 33 and 34 are truncated normal densities and distributions, respectively. Since the \( u_i \) are iid in order to compute the probabilities the modeler has to substitute the MLE of the respective parameters into the conditional distribution \( (f(u|\varepsilon)) \) in equation 32 and then compute the probabilities. The presence of spatial correlation in the \( v_i \) induces an idiosyncratic variance-covariance matrix with heteroskedasticity and correlation, but this does not affect parameters of the inefficiency distribution - the iid assumption is preserved. Equation 33 represents the probability that firm \( k \) is drawing a value of \( u \) that is larger than all the other firms in the population which means that firm \( k \) may deem the least efficient while the converse holds true for equation 34 in which firm \( k \) is deemed the most efficient. Equations 33 and 34 hold for any general absolutely continuous pdfs and cdfs.

**Result 8** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \delta \overset{\text{iid}}{\sim} \left| N(0, \sigma^2_{u}I_n) \right| \), the probability that firm \( k \) is the least and most efficient is computed by using the product of an unknown, non-standard marginal density and univariate conditional cdfs, see equations 35 and 36.

\[ P(u_j \leq u_k, \ j \neq k = 1) = \int_0^\infty f_{u_1}(\mu_1, \sigma_{u_1}; u_1) \prod_{j \neq k}^{n-2} F_{u_{j\mid u_{j-1}}(\mu_j, \sigma_{u_{j\mid u_{j-1}}}; u_k)} du_{j\neq k=1} \]  
(35)

\[ P(u_j > u_k, \ j \neq k = 1) = \int_0^\infty f_{u_1}(\mu_1, \sigma_{u_1}; u_1) \prod_{j \neq k}^{n-2} (1 - F_{u_{j\mid u_{j-1}}(\mu_j, \sigma_{u_{j\mid u_{j-1}}}; u_k)}) du_{j\neq k=1} \]  
(36)
Note that \( f_{u_1}(\mu_1, \sigma_{u_1}; u_1) \) and \( F_{u_j|u_{j-1}} \) are unknown non-standard density and conditional distribution function, respectively. Result 8 uses SC to reduce the joint distribution into the product of an unknown marginal density and conditional densities. In order to use result 8 to compute the probabilities it requires information about the joint distribution, that is, the modeler has to compute the marginal density by integrating \( n-1 \) integrals. MLEs of the parameters are substituted directly into the conditional distributions into equations 35 and 36, respectively.

**Result 9** If \( \rho \neq 0 \) and \( \eta = 0 \), and \( \delta \stackrel{iid}{\sim} N(0, \sigma_u^2 I_n) \), the probability that firm \( k \) is the least and most efficient is computed by using the product of a marginal truncated normal density and univariate conditional truncated normal cdfs, see equations 37 and 38.

\[
P(u_j \leq u_k; j \neq k = 1) = \int_0^\infty f_{u_1}(\mu_1, \sigma_{u_1}; u_1) \prod_{j=2}^n (F_{u_j|u_{j-1}}(\mu_j, \sigma_{u_j}; u_k) du_{j \neq k=1} = 1)
\]

\[
P(u_j > u_k; j \neq k = 1) = \int_0^\infty f_{u_1}(\mu_1, \sigma_{u_1}; u_1) \prod_{j=2}^n (1 - F_{u_j|u_{j-1}}(\mu_j, \sigma_{u_j}; u_k) du_{j \neq k=1} = 1)
\]

The appendix provides the proof. \( F_{u_j|u_{j-1}} \) is a univariate conditional truncated normal cdf of \( u_j \) conditioned on \( u_{j-1} \) and \( f_1(\mu_1, \sigma_{u_1}; u_1) \) is a truncated normal density. While results 8 and 9 use SC to factor the joint distribution into the product of conditional densities and a marginal density, result 9 yields a standard, known density and distributions. This is relatively easier than result 8 in calculating the probabilities. The marginal density using result 9 has a standard analytical form therefore integrating \( n-1 \) integrals is an unnecessary step. Using SC for result 2 the numerator and denominator are factored in normal random variables. A linear combinations of normal random variables are normally distributed. As such the multivariate truncated normal is factored into a univariate marginal truncated normal and univariate conditional truncated normal densities. Multiple integrals are difficult to solve therefore using the univariate conditional distributions
instead of the multivariate distribution simplifies the computation of the probabilities.

The mean of the univariate conditional densities/distributions will be a function of the realizations of the random variables being conditioned on, such that \(u_j|u_{j-1} = \hat{u}_{j-1}\), where \(\hat{u}_{j-1}\), \(j=2,...,n\) are the realizations of the random variables. The conditional mean function \(\bar{E}(u|\varepsilon_{j-1} = e_i) = \hat{u}_{j-1}\) is used to calculate inefficiency for each firm \(i\) which are then substituted into the mean of the conditional distributions.

Under normality assumption let \(u_1\) and \(u_2\) be \(n_1 \times 1\) and \(n_2 \times 1\) subvectors of the \(n \times 1\), random vector \(u\) such that \((n_1 + n_2 = n)\), all the conditional mean and variances are computed using the standard formula:

\[
u_2|u_1 \sim N(\Sigma_{\sigma_{u21}}\Sigma_{u11}^{-1} u_1, \Sigma_{\sigma_{u22}} - \Sigma_{\sigma_{u21}}\Sigma_{u11}^{-1}\Sigma_{\sigma_{u12}})\]  

where \(\Sigma_{\sigma_{u21}} = \Sigma_{\sigma_{u12}}\) are the covariances and \(\Sigma_{\sigma_{u11}}\) and \(\Sigma_{\sigma_{u22}}\) are the variances between \(u_1\) and \(u_2\), respectively. The estimate of the variance-covariance matrix of inefficiencies are recovered by substituting MLE of \((\sigma_u^2\) and \(\rho\) which are \((\hat{\sigma}_u^2\) and \(\hat{\rho}\)) into \(\hat{\Sigma}_\sigma\). We can also recover \(\hat{\Sigma}_{\sigma,\sigma}\) = \(\hat{\Sigma}_\sigma + \frac{l}{\hat{\sigma}_u^2}\). From the estimate of \(\hat{\Sigma}_\sigma\) we will be able to know estimates for \((\Sigma_{\sigma_{u21}}\), \(\Sigma_{\sigma_{u12}}\), \(\Sigma_{\sigma_{u11}}\) and \(\Sigma_{\sigma_{u22}}\) and then proceed to compute the probabilities. Let \(\sigma_{ij}\) for \((i=1,...,n\) and \(j=1,...,n)\) be the element of \(\Sigma_{\sigma}\) in which \(i = j\) represents variances and \(i \neq j\) represents covariances, and \(\hat{\sigma}_{ij}\) are the elements of \(\hat{\Sigma}_\sigma\) and these are used in the computation of result 9.\(^{20}\)

Using result 9, suppose \(n = 3\), the probability that firm \(k\) is the least efficient is given by:\(^{21}\)

\[
P(u_3 \leq u_1, u_2 \leq u_1) = \int_0^\infty \frac{f_{u_1}(\mu_1, \sigma_{u1}; u_1)}{(1 - F_{u_1}(\mu_1, \sigma_{u1}))} \frac{F_{u_2|u_1}(\mu_2, \sigma_{u2}; u_2)}{(1 - F_{u_2|u_1}(\mu_2, \sigma_{u2}))} du_1 \]  

\[
P(u_1 \leq u_2, u_3 \leq u_2) = \int_0^\infty \frac{f_{u_2|u_1}(\mu_2, \sigma_{u2}; u_2)}{(1 - F_{u_2|u_1}(\mu_2, \sigma_{u2}))} \frac{F_{u_3|u_2, u_1}(\mu_3, \sigma_{u3}; u_3)}{(1 - F_{u_3|u_2, u_1}(\mu_3, \sigma_{u3}))} du_2 \]  

\[
P(u_1 \leq u_3, u_2 \leq u_3) = \int_0^\infty \frac{f_{u_3|u_2, u_1}(\mu_3, \sigma_{u3}; u_3)}{(1 - F_{u_3|u_2, u_1}(\mu_3, \sigma_{u3}))} \frac{F_{u_1|u_3, u_2, u_1}(\mu_1, \sigma_{u1}; u_1)}{(1 - F_{u_1|u_3, u_2, u_1}(\mu_1, \sigma_{u1}))} du_3 \]  

\(^{20}\)In software packages such as STATA and Matlab, it is straight forward to substitute the estimates of \((\hat{\sigma}_u^2\) and \(\hat{\rho}\)) into \(\hat{\Sigma}_\sigma\) and compute the variance-covariance matrix of inefficiency.

\(^{21}\)See Appendix for when \(n=2\).
while the probability that firm $k$ is the most efficient is:

$$P(u_3 > u_1, u_2 > u_1) = \int_0^\infty \frac{f_{u_1}(\mu_1, \sigma_{u_1}; u_1)}{(1 - F_{u_1}(\mu_1, \sigma_{u_1}))} \frac{(1 - F_{u_3|u_1}(\mu_3, \sigma_{u_3}; u_1))}{(1 - F_{u_3|u_1}(\mu_3, \sigma_{u_3}))} du_1$$

$$P(u_1 > u_2, u_3 > u_2) = \int_0^\infty \frac{f_{u_2|u_1}(\mu_2, \sigma_{u_2}; u_2)}{(1 - F_{u_2|u_1}(\mu_2, \sigma_{u_2}))} \frac{(1 - F_{u_3|u_2}(\mu_3, \sigma_{u_3}; u_2))}{(1 - F_{u_3|u_2}(\mu_3, \sigma_{u_3}))} du_2$$

$$P(u_1 > u_3, u_2 > u_3) = \int_0^\infty \frac{f_{u_3|u_2,u_1}(\mu_3, \sigma_{u_3}; u_3)}{(1 - F_{u_3|u_2,u_1}(\mu_3, \sigma_{u_3}))} \frac{(1 - F_{u_1|u_2}(\mu_1, \sigma_{u_1}; u_1))}{(1 - F_{u_1|u_2}(\mu_1, \sigma_{u_1}))} du_3$$

For a row normalize weighting matrix, $0 < w_{ij} \leq 1$ for $i \neq j$. Let $a_{ij}(i = 1, 2 \text{ and } 3 \text{ and } j = 1, 2 \text{ and } 3)$ be the elements of the matrix $A$. Let $D$ be the determinant of $A$ and $\sigma_{ij} = \frac{\sigma^2_{a_{ij}}}{D^2}$. The $a_{ij}$ are defined in the appendix. The distribution $f_{u_1}(u_1)$ is truncated from the normal density:

$$u_1^* \sim N(0, \frac{\sigma^2_{a_{11}}}{D^2})$$

the conditional distribution of $f_{u_2|u_1}(u_2|u_1 = \tilde{u}_1)$ where $\tilde{u}_1 = \tilde{E}(u_1|\varepsilon_1 = e_1)$ and $(e_1$ is the residuals) is truncated from the normal density:

$$u_2^*|u_1^* \sim N(\frac{a_{21}\tilde{u}_1}{a_{11}}, \frac{\sigma^2_{a_{11}}(a_{22}a_{11} - a^2_{12})}{D^2a_{11}})$$

and $f_{u_3|u_2,u_1}(u_3|u_2 = \tilde{u}_2, u_1 = \tilde{u}_1)$ where $\tilde{u}_1 = \tilde{E}(u_1|\varepsilon_1 = e_1)$ and $\tilde{u}_2 = \tilde{E}(u_2|\varepsilon_2 = e_2)$ is truncated from the normal density:

$$u_3^*|u_2^* \sim N((a_{31}a_{22} - a_{32}a_{21})\tilde{u}_1 + (\frac{a_{32}a_{11} - a_{31}a_{12}}{a_{11}a_{22} - a^2_{12}})\tilde{u}_2), \frac{\sigma^2_{a_{11}}((a_{31}a_{22} - a_{32}a_{21})\tilde{u}_1 + (\frac{a_{32}a_{11} - a_{31}a_{12}}{a_{11}a_{22} - a^2_{12}})\tilde{u}_2)^2}{D^2(a_{33} - (\frac{a_{31}a_{22} - a_{32}a_{21}}{a_{11}a_{22} - a^2_{12}}))}$$

where $u_2^* = [u_2^* \ u_1^*']$ is a 2 $\times$ 1 vector. If the spatial parameter $(\rho)$ equals zero then $D = 1$ and $a_{ij} = 0$ for
\( i \neq j \), and \( a_{ij} = 1 \) for \( i = j \). Results 7 to 9 become computationally involve as the number of firms increases. Result 9 has known standard pdfs and cdfs with just one integral so the probabilities can be simulated when the sample size becomes relatively large.

7 Conclusion

This paper extends the current literature by adding spatial correlation to the production function for a cross section of firms and generalizing Horrace (2005) probability statements to account for this cross-sectional dependence. The presence of spatial correlation results in an intractable likelihood, since the number of integrals increases with the sample size. This paper assumes that inefficiency is drawn from a multivariate truncated normal distribution with zero mean. However, it could also be assumed that inefficiency is drawn from a multivariate truncated normal distribution with a non-zero mean prior to truncation, a multivariate gamma distribution or a multivariate doubly truncated normal distribution.\(^{22}\) A Monte carlo study of the likelihood is limited because the number of integrals is restricted to only twenty five.

COLS estimates are consistent and can be substituted into the probability statements. However, these estimates are less efficient when compared to MLES, so making inferences will be a concern. The technique of SC eases the calculation of the probability of the least and most efficient firm because it allows the joint distribution to factor into the product of a univariate marginal distribution and univariate conditional distributions which are computationally less involved when calculating the probabilities. Unlike Horrace (2005) if the spatial correlation is in the inefficiency distribution the MLEs or COLS are substituted directly into the probability statements instead of inefficiency conditioned on the composed error. However, if the spatial correlation is in the idiosyncratic component \((v)\) the probability statements are reduced to Horrace (2005) equations 4 and 5, however the variance-covariance matrix has a correlation structure that must be

\(^{22}\)Steven (1990) generalizes the univariate half normal to the univariate truncated normal with a non-zero mean and the exponential to the gamma distribution. Almanidis, Qian and Sickles (2014) generalize the univariate truncated normal to a doubly truncated normal.
accounted for in order to make more reliable inferences.

Horrace, Richards-Shubik and Wright (2015) state that the conditional distributions of inefficiency are based on the errors which are independent even though OLS residuals are correlated, and this might be a concern when addressing inference. The technique of SC provides some insight on how to deal with non-random samples. Unlike the errors which are unobservable, the residuals are correlated and known, so SC can provide some guidance on how to proceed. Future research should explore generalizing the expected efficiency rank by Horrace, Richards-Shubik and Wright (2015) so that one can determines the best or worst firm in the sample when the error term is dependent.

Spatial correlation induces heteroskedascity that will distort the firms’ ranking. The traditional conditional mean function will produce bias estimates for inefficiency while Horrace (2005) probability statements are incorrect and unreliable. Given parametric assumptions this paper has provided some insights for empiricists on how to proceed in computing point estimates of inefficiency as well as calculating the probability statements in the presence of a specific cross-sectional dependence.

References


Appendix: Proofs of results

This appendix provides all the proofs for the results in the text.

A Proof of Result 4: Start with a truncated normal before spatial correlation is added

\[
\begin{align*}
    u^* &= \rho W_1 u^* + \delta \quad \text{and} \quad \delta \overset{iid}{\sim} N(0, I_n \sigma_u^2) \quad (49) \\
    u^* &= (I_n - \rho W_1)^{-1} \delta \Rightarrow u^* \sim (0, \Omega_{\sigma_u}) \quad (50) \\
    \Omega_{\sigma_u} &= \sigma_u^2 (I_n - \rho W_1)' (I_n - \rho W_1)^{-1} \text{ is a symmetric and positive definite matrix}
\end{align*}
\]
Note that \( u = u^* \in [0, \infty) \) is a \( n \times 1 \) vector, the distribution of the \( u \) is:

\[
f(u) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_n} \right)^n \exp\left( -\frac{1}{2} (u' \Omega_{\sigma_n}^{-1} u) \right) |J| \tag{51}
\]

where \( |J| = |I_n - \rho W_1| \) is the determinant of the Jacobian

\[
\frac{|J|}{\sigma_n^n} = |\Omega_{\sigma_n}|^{-\frac{1}{2}}
\]

\[
\Rightarrow f(u) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \frac{1}{(\det \Omega_{\sigma_n})^{\frac{1}{2}}} \exp\left( -\frac{1}{2} (u' \Omega_{\sigma_n} u) \right) \tag{52}
\]

which is not a multivariate normal distribution nor a multivariate truncated normal distribution.

\[
f(v) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_v} \right)^n \exp\left( -\frac{1}{2\sigma_v^2} (v' v) \right) \tag{53}
\]

Impose \( A\beta \) and note that \( \varepsilon + u = v \) with the Jacobian equals to 1. The joint distribution of \( \varepsilon \) and \( u \) is:

\[
f(u, \varepsilon) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_v} \right)^n \frac{1}{(\det \Omega_{\sigma_v})^{\frac{1}{2}}} \exp\left( -\frac{1}{2\sigma_v^2} (\varepsilon + u)' (\varepsilon + u) \right) \exp\left( -\frac{1}{2} (u' \Omega_{\sigma_v}^{-1} u) \right) \tag{54}
\]

The exponent simplifies to:

\[
-\frac{1}{2\sigma_v^2} (\varepsilon \varepsilon + 2\varepsilon' u + u' u) - \frac{1}{2} (u' \Omega_{\sigma_v}^{-1} u)
\]

\[
-\frac{1}{2\sigma_v^2} (\varepsilon \varepsilon + 2\varepsilon' u) - \frac{1}{2} (u' \Omega_{\sigma_v}^{-1} u) \tag{55}
\]

where \( \Omega_{\sigma_v \sigma_u}^{-1} = I_n \frac{n}{\sigma_v^2} + \Omega_{\sigma_u}^{-1} \) is a symmetric and positive definite matrix.
completing the square:

\[- \frac{1}{2} \left( \frac{2 \varepsilon' u}{\sigma_v^2} + u \Omega_{\sigma_v \sigma_u}^{-1} u \right) \]

add and subtract this term \( \frac{\varepsilon'}{\sigma_v^2} \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2} \)

\(- \frac{1}{2 \sigma_v^2} (\varepsilon' \varepsilon) + \frac{1}{2} \frac{\varepsilon' \varepsilon}{\sigma_v^2} \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2} - \frac{1}{2} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \Omega_{\sigma_v \sigma_u}^{-1} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \)

\[ f(u, \varepsilon) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_v} \right)^n \exp \left( \frac{1}{2} \left( - \frac{\varepsilon' \varepsilon}{\sigma_v^2} + \frac{\varepsilon' \varepsilon}{\sigma_v^2} \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2} \right) \right) \frac{1}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \right) \]

\[ f(\varepsilon) = Z_3 \left( \frac{1}{\sigma_v} \right)^n \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{\det \Omega_{\sigma_v \sigma_u}}{\det \Omega_{\sigma_v \sigma_u}} \right)^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \right) \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \right) \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} (u + \Omega_{\sigma_v \sigma_u} \frac{\varepsilon}{\sigma_v^2}) \right) \]

The composed error is intractable since \( Z_3 \) has integrals that increases with the sample size.
Proof of Result 5: Multivariate Truncated Normal for Inefficiency

\[ u^* = \rho W_1 u^* + \delta \text{ and } \delta \overset{iid}{\sim} N(0, I_n \sigma_u^2) \]  
(60)

\[ u^* = (I_n - \rho W_1)^{-1} \delta \Rightarrow u^* \sim N(0, \Sigma_{\sigma_u}) \]  
(61)

\[ \Sigma_{\sigma_u} = \sigma_u^2 (I_n - \rho W_1)' (I_n - \rho W_1)^{-1} \text{ is a symmetric and positive definite matrix} \]

Let \( u \in [0, \infty) \) be a \( n \times 1 \) random vector truncated from \( u^* \), the distribution of \( u \) is:

\[ f(u) = \frac{f(u^*)}{\Pr(u_1^* > 0, u_2^* > 0, \ldots, u_n^* > 0)} \]  
(62)

\[ f(u) = \frac{(\frac{1}{\sqrt{2\pi}})^n \det(\Sigma_{\sigma_u})^{-\frac{1}{2}} \exp(-\frac{1}{2}(u' \Sigma_{\sigma_u}^{-1} u))}{Q_0} \]  
(63)

where \( Q_0 = \int_0^\infty \cdots \int_0^\infty \left( \frac{1}{\sqrt{2\pi}} \right)^n \det(\Sigma_{\sigma_u})^{-\frac{1}{2}} \exp(-\frac{1}{2}(u' \Sigma_{\sigma_u}^{-1} u)) du_1, \ldots, du_n \)

Let \( v \in (-\infty, \infty) \) be a \( n \times 1 \) vector which is iid and \( N(0, I_n \sigma_v^2) \) such that:

\[ f(v) = (\frac{1}{\sqrt{2\pi}})^n \left( \frac{1}{\sigma_v} \right)^n \exp(-\frac{1}{2\sigma_v^2}(v' v)) \]  
(64)

Impose \( A3 \) and note that \( \varepsilon + u = v \) with the Jacobian equals to 1. The joint distribution of \( \varepsilon \) and \( u \) is:

\[ f(u, \varepsilon) = \frac{1}{\sqrt{2\pi}}^n \left( \frac{1}{\sigma_v} \right)^n \frac{1}{Q_0} \exp(-\frac{1}{2\sigma_v^2}(\varepsilon + u)' (\varepsilon + u)) \exp(-\frac{1}{2}(u' \Sigma_{\sigma_u}^{-1} u)) \]  
(65)
The exponent simplifies to:

\[- \frac{1}{2 \sigma_v^2} (\varepsilon \varepsilon + 2 \varepsilon \cdot u + u \cdot u) - \frac{1}{2} (u \cdot \Sigma_{\sigma_u}^{-1} u) \]

\[- \frac{1}{2 \sigma_v^2} (\varepsilon \varepsilon + 2 \varepsilon \cdot u) - \frac{1}{2} (u \cdot \Sigma_{\sigma_u}^{-1} u) \tag{66} \]

where \( \Sigma_{\sigma_u}^{-1} = \frac{I_n}{\sigma_v^2} + \Sigma_{\sigma_u}^{-1} \) is a symmetric and positive definite matrix

completing the square:

\[- \frac{1}{2} \left( \frac{2 \varepsilon \cdot u}{\sigma_v^2} + u \cdot \Sigma_{\sigma_u}^{-1} u \right) \]

\[- \frac{1}{2} \left( \frac{2 \varepsilon \cdot u}{\sigma_v^2} + u \cdot \Sigma_{\sigma_u}^{-1} u \right) \tag{67} \]

add and subtract this term \( \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} \)

\( \Rightarrow: \)

\[- \frac{1}{2 \sigma_v^2} (\varepsilon \varepsilon) + \frac{1}{2} \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} - \frac{1}{2} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \cdot \Sigma_{\sigma_u}^{-1} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \]

\[ f(u, \varepsilon) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\det \Sigma_{\sigma_u}} \right)^{\frac{1}{2}} \frac{1}{Q_0} \exp \left( \frac{1}{2} \left( \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} + \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} \right) \right) \exp \left( - \frac{1}{2} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \cdot \Sigma_{\sigma_u}^{-1} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \right) \tag{68} \]

\[ f(u, \varepsilon) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\det \Sigma_{\sigma_u}} \right)^{\frac{1}{2}} \frac{1}{Q_0} \exp \left( \frac{1}{2} \left( \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} + \frac{\varepsilon \cdot \Sigma_{\sigma_u}^{-1} \varepsilon}{\sigma_v^2} \right) \right) \exp \left( - \frac{1}{2} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \cdot \Sigma_{\sigma_u}^{-1} \left( u + \Sigma_{\sigma_u}^{-1} \varepsilon \right) \right) \tag{69} \]
The composed error has a distribution $Z_0$ and $Q_0$ with integrals that increases with the sample size. The loglikelihood of the $n \times 1$ vector $y$ is:

$$\ln(\beta, \sigma^2, \sigma^2_v, \rho; y) = -n \ln(-\sqrt{2\pi}) - n \ln(\sigma_v) + \ln(Z_0(y - x\beta)) - \ln(Q_0) + 0.5 \ln(\det \Sigma_{v,\sigma})$$

$$-0.5 \ln(\det \Sigma_{\sigma}) + \frac{1}{2} \left( \frac{(y - x\beta)'(y - x\beta)}{\sigma_v^2} + \frac{(y - x\beta)'\Sigma_{v,\sigma}(y - x\beta)}{\sigma_v^4} \right)$$

**C Proof of Result 6: Spatial correlation in the $v$**

If the spatial correlation is in the $v$ then:
\[ v = \eta W_2 v + \zeta \]  
(73)

\[ v = (I_n - \eta W_2)^{-1} \zeta \text{ and } \zeta \overset{\text{iid}}{\sim} N(0, I_n \sigma^2_v) \]  
(74)

\[ \Rightarrow v \sim N(0, \Sigma_{\sigma_v}) \]  
(75)

\[ \Sigma_{\sigma_v} = \sigma^2_v ((I_n - \eta W_2)' (I_n - \eta W_2))^{-1} \] is a symmetric and positive definite matrix

The \( u \overset{\text{iid}}{\sim} N(0, I_n \sigma^2_u) \)

\[ f(u) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_u} \right)^n \exp\left( -\frac{1}{2} \frac{u'u}{\sigma_u^2} \right) \]  
(76)

\[ f(v) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{(\det \Sigma_{\sigma_v})^{\frac{1}{2}}} \right) \exp\left( -\frac{1}{2} v'^{-1}v \right) \]  
(77)

completing the square:

\[ \pm \varepsilon^\prime \Sigma_{\sigma_v}^{-1} \Sigma_{\sigma_v u}^{* \prime \sigma_u} \Sigma_{\sigma_v}^{-1} \varepsilon \]

and let \( \Sigma_{\sigma_v u}^{* \prime \sigma_u} = \Sigma_{\sigma_v}^{-1} + \frac{I_n}{\sigma_u^2} \) is a symmetric positive definite matrix  
(78)

\[ f(u, \varepsilon) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\det(\Sigma_{\sigma_v})^{\frac{1}{2}}} \right) \left( \frac{1}{\sigma_u} \right)^n \exp\left( \frac{1}{2} \left( -\varepsilon^\prime \Sigma_{\sigma_v}^{-1} \varepsilon + \varepsilon^\prime \Sigma_{\sigma_v}^{-1} \Sigma_{\sigma_v u}^{* \prime \sigma_u} \Sigma_{\sigma_v}^{-1} \varepsilon \right) \right) \]

\[ \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp\left( -\frac{1}{2} (u + \Sigma_{\sigma_v u}^{* \prime \sigma_u} \Sigma_{\sigma_v}^{-1} \varepsilon)'^\prime \Sigma_{\sigma_v u} \Sigma_{\sigma_v}^{-1} \varepsilon \right) \]  
(79)
After some algebra:

\[ f(\varepsilon) = \left( \frac{2}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\sigma_u} \right)^n \left( \frac{1}{\sigma_v} \right)^n \left( \frac{1}{\det \Sigma_{\sigma_u\sigma_u}} \right)^{\frac{1}{2}} A_0 \exp \left( \frac{1}{2} \left( -\varepsilon \Sigma_{\sigma_u \sigma_u}^{-1} \varepsilon' + \varepsilon' \Sigma_{\sigma_v \sigma_v}^{-1} \Sigma_{\sigma_v \sigma_u}^{-1} \Sigma_{\sigma_u \sigma_v} \varepsilon \right) \right) \]

\[ A_0 = \Pr \left( u_1 \geq 0, \ldots, u_n \geq 0 \right) = \int_0^\infty \cdots \int_0^\infty \left( \frac{1}{\sqrt{2\pi}} \right)^n \left( \frac{1}{\det \Sigma_{\sigma_v \sigma_u}} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \left( u + \Sigma_{\sigma_v \sigma_u} \Sigma_{\sigma_u \sigma_v}^{-1} \varepsilon \right)' \Sigma_{\sigma_v \sigma_u}^{-1} (u + \Sigma_{\sigma_v \sigma_u} \Sigma_{\sigma_u \sigma_v}^{-1} \varepsilon) \right) \right) \]

\[ \ln(\beta, \sigma_u^2, \sigma_v^2, \eta; y) = n \ln(2) - n \ln(\sqrt{2\pi}) - n \ln(\sigma_u) - 0.5 \ln(\det \Sigma_{\sigma_u}) + \ln(A_0) + 0.5 \ln(\det \Sigma_{\sigma_v \sigma_v}) + \frac{1}{2} \left( -(y - X\beta) \Sigma_{\sigma_v}^{-1} \Sigma_{\sigma_v \sigma_u}^{-1} \Sigma_{\sigma_u \sigma_v} (y - X\beta) \right) \]

\[ (81) \]

\[ A_0 \] has integrals that increase with the sample size.

### D Spatial correlation in the both v and u

If the spatial correlation is in both \( u \) and \( v \) and \( W_1 \neq W_2 \) then:
\[ u^* = \rho W_1 u^* + \delta \quad (82) \]
\[ v = \eta W_2 v + \zeta \quad (83) \]
\[ \delta \overset{iid}{\sim} N(0, I_n \sigma_u^2) \]
\[ \zeta \overset{iid}{\sim} N(0, I_n \sigma_v^2) \]

\[ \Rightarrow u^* \sim N(0, \Sigma_{\sigma_u}) \]
\[ \Rightarrow v \sim N(0, \Sigma_{\sigma_v}) \]

\[ \Sigma_{\sigma_u} = \sigma_u^2 ((I_n - \rho W_1)(I_n - \rho W_1))^{-1} \]
\[ \Sigma_{\sigma_v} = \sigma_v^2 ((I_n - \eta W_2)(I_n - \eta W_2))^{-1} \]

\[ f(u) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \frac{1}{(\det \Sigma_{\sigma_u})^{\frac{1}{2}}} \frac{1}{Q_0} \exp \left( -\frac{1}{2} (u^T \Sigma_{\sigma_u}^{-1} u) \right) \]
\[ f(v) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \frac{1}{(\det \Sigma_{\sigma_v})^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (v^T \Sigma_{\sigma_v}^{-1} v) \right) \]

After some algebra:

\[ f(\varepsilon) = \frac{(\det \Sigma_{\sigma_u,\sigma_v})^{\frac{1}{2}}}{(\det \Sigma_{\sigma_v})^{\frac{1}{2}} (\det \Sigma_{\sigma_u})^{\frac{1}{2}}} \frac{F_0}{Q_0} \exp \left( -\frac{1}{2} \varepsilon^T \Sigma_{\sigma_u,\sigma_v}^{-1} \varepsilon + \frac{1}{2} \varepsilon^T \Sigma_{\sigma_u,\sigma_v}^{-1} \Sigma_{\sigma_v}^{-1} \varepsilon \right) \]
\[ F_0 = \Pr (u_1 \geq 0, \ldots, u_n \geq 0) = \int_0^{\infty} \cdots \int_0^{\infty} \left( \frac{1}{\sqrt{2\pi}} \right)^n \frac{1}{(\det \Sigma_{\sigma_u,\sigma_v})^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (u^T \Sigma_{\sigma_u,\sigma_v}^{-1} u) \right) \]
\[ \Sigma_{\sigma_u,\sigma_v} = \Sigma_{\sigma_u} + \Sigma_{\sigma_v} \]
\[ \Sigma_{\sigma_u,\sigma_v}^{-1} \text{ is a symmetric, positive definite matrix} \]

\[ F_0 \text{ and } Q_0 \text{ have integrals that increase with the sample size.} \]
E  The Marginal Distribution using Result 1

For a truncated normal the marginal distribution is:

\[
\begin{align*}
    f_{u_1} &= \int_{0}^{\infty} \frac{f_{u_1 u_2}()}{Q_0} du_2 \\
    f_{u_1} &= \frac{1}{\sqrt{2\pi} \sigma_u} \frac{1}{(1 + \rho^2 w_{12}^2)^\frac{1}{2}} \frac{1}{Q_0} \left(1 - \Phi\left(-\frac{u_1 \rho(w_{12} + w_{21})}{\sigma_u(1 + \rho^2 w_{12}^2)^\frac{1}{2}}\right) \exp\left(-\frac{u_1^2 (1 - \rho^2 w_{21} w_{12})^2}{2\sigma_u^2 (1 + \rho^2 w_{12}^2)}\right)\right)
\end{align*}
\]

For \( n = 2 \) suppose the assumption on \( \delta \) is and note that \( u^* = u \geq 0 \)

\[
    u^* = W_1 u^* + \delta \quad \text{and} \quad \delta \sim N(0, I_n \sigma_u^2)
\]

\[
    f_{\delta_1 \delta_2}() = \frac{1}{2\pi \sigma_u} \exp\left(-\frac{1}{2\sigma_u^2} (\delta_1^2 + \delta_2^2)\right)
\]

\[
    \delta = [I_n - \rho W_1] u^*
\]

F  The Marginal Distribution using Result 2

For \( n = 2 \),

\[
    u^* = W_1 u^* + \delta \quad \text{and} \quad \delta \sim N(0, I_n \sigma_u^2)
\]

\[
    u^* \sim N(0, \Sigma_{\sigma_u})
\]
\[
\delta_1 = \begin{bmatrix} 1 & -\rho w_{12} \\ -\rho w_{21} & 1 \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} \\
\delta_2 = \begin{bmatrix} u_1^* - \rho w_{12} u_2^* \\ -\rho w_{21} u_1^* + u_2^* \end{bmatrix}
\]

\[
\begin{align*}
\delta_1 & = u_1^* - \rho w_{12} u_2^* \\
\delta_2 & = -\rho w_{21} u_1^* + u_2^* \\
|J| & = (1 - \rho^2 w_{21} w_{12}) \\
\end{align*}
\]

After some algebra:

\[
f_{u_1 u_2}(\cdot) = \frac{1}{2\pi} \frac{|J|}{\sigma_u^2} \exp\left(-\frac{1}{2\sigma_u^2} ((u_1^* - \rho w_{12} u_2^*)^2 + (-\rho w_{21} u_1^* + u_2^*)^2) \right) 
\]

\[
f_{u_1} = \int_{-\infty}^{\infty} f_{u_1 u_2}(\cdot) du_2
\]

\[
f_{u_1}(u_1) = \frac{1}{\sqrt{2\pi} \sigma_u} \frac{1}{(1 + \rho^2 w_{12}^2)^{\frac{1}{2}}} \exp\left(-\frac{u_1^2 (1 - \rho^2 w_{21} w_{12}^2)}{2\sigma_u^2 (1 + \rho^2 w_{12}^2)} \right)
\]

\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_u} \frac{1}{(1 + \rho^2 w_{12}^2)^{\frac{1}{2}}} \exp\left(-\frac{(1 + \rho^2 w_{12}^2)}{2\sigma_u^2} (u_2^* - A)^2 \right) du_2^*
\]

\[
A = \frac{u_1^* \rho (w_{12} + w_{21})}{(1 + \rho^2 w_{12}^2)}
\]

\[
f_{u_1}(u_1) = \frac{1}{\sqrt{2\pi} \sigma_u} \frac{1}{(1 + \rho^2 w_{12}^2)^{\frac{1}{2}}} \exp\left(-\frac{u_1^2 (1 - \rho^2 w_{21} w_{12}^2)^2}{2\sigma_u^2 (1 + \rho^2 w_{12}^2)} \right)
\]
G Sequential Conditioning

The joint distribution is:

\[ f(u_1, \ldots, u_n) = \frac{f(u_1, \ldots, u_n)}{Pr(u_1 \geq 0, \ldots, u_n \geq 0)} \]

The paper uses SC, see Spanos (1986;1999) to reduce the joint distribution into a single marginal distribution and the product of conditional distributions, here \( \Psi_j \) represents the parameter space associated with the respective distributions.

\[
\begin{align*}
    f(u_1, \ldots, u_n) &= \frac{f(u_j | u_{j-1}, \ldots, u_1; \Psi_j) \times f(u_{j-1} | u_{j-2}, \ldots, u_1; \Psi_{j-1}) \times \cdots \times f(u_1; \Psi_1)}{\int_0^\infty \cdots \int_0^\infty f(u_j | u_{j-1}, \ldots, u_1; \Psi_j) \times f(u_{j-1} | u_{j-2}, \ldots, u_1; \Psi_{j-1}) \times \cdots \times f(u_1; \Psi_1) du_1 \cdots du_n} \\
    f(u_1, \ldots, u_n) &= \frac{f(u_1) \prod_{j=2}^n f(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j)}{\int_0^\infty \cdots \int_0^\infty f(u_1; \Psi_1) \prod_{j=2}^n f(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j) du_1 \cdots du_n} \\
    f(u_1, \ldots, u_n) &= \frac{f(u_1) \prod_{j=2}^n f(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j)}{\int_0^\infty \cdots \int_0^\infty f(u_1; \Psi_1) \prod_{j=2}^n f(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j) du_1 \cdots du_n} \\
    f(u_1, \ldots, u_n) &= \frac{f(u_1) \prod_{j=2}^n f(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j)}{(1 - F(u_1; \Psi_1)) \prod_{j=2}^n (1 - F(u_j | u_{j-1}, u_{j-2}, \ldots, u_1; \Psi_j)) du_1 \cdots du_n} \\
\end{align*}
\]

(103)

given the assumption of normality on \( u \) the joint distribution is reduced the product of a marginal truncated normal multiply by univariate conditional truncated normal distributions.
H Spatial Correlation for n=3

Using result 2:

\[
\begin{align*}
\text{u}^* & \sim N(0, \Sigma_{\sigma_u}) \\
E(\text{u}^*) & = [I_n - \rho W_n]^{-1}E(\delta) = 0 \\
\Sigma_{\sigma_u} & = \sigma_u^2([I_n - \rho W_n]'(I_n - \rho W_n)]^{-1}
\end{align*}
\]

recall that \(u^*\) is \(n \times 1\) vector and let \(u_1^*\) and \(u_2^*\) be \(n_1 \times 1\) and \(n_2 \times 1\) subvectors, respectively. Let \(\Sigma_{\sigma_u} = \sigma_u^2[A'A]^{-1}\) where \(A = (I_n - \rho W_n)\), using result 2 the conditional distribution prior to truncation is:

\[
u_2^* | u_1^* \sim N(\Sigma_{\sigma_{u_{21}}}^{-1} u_1, \Sigma_{\sigma_{u_{22}}} - \Sigma_{\sigma_{u_{21}}} \Sigma_{\sigma_{u_{11}}} \Sigma_{\sigma_{u_{12}}}) (104)\]

For \(n = 2\), the variance-covariance matrix has components, \(\sigma_{11}, \sigma_{22}, \sigma_{12}\) and \(\sigma_{21}\) which are the variance of \(u_1^*, u_2^*\) and covariance of \((u_1, u_2)\), respectively.

\[
\begin{align*}
\Sigma_{\sigma_u} & = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \\
A & = (I_n - \rho W_n) = \begin{bmatrix} 1 & -\rho w_{12} \\ -\rho w_{21} & 1 \end{bmatrix} \\
\Sigma_{\sigma_u} & = \sigma_u^2[A'A]^{-1} = \begin{bmatrix} \frac{\sigma^2 (1+\rho^2 w_{12}^2)}{1-\rho^2 w_{12} w_{21}} & \frac{\sigma^2 \rho (w_{12} + w_{21})}{1-\rho^2 w_{12} w_{21}} \\ \frac{\sigma^2 \rho (w_{12} + w_{21})}{1-\rho^2 w_{12} w_{21}} & \frac{\sigma^2 (1+\rho^2 w_{21}^2)}{1-\rho^2 w_{12} w_{21}} \end{bmatrix} (105)
\end{align*}
\]
\[
\begin{align*}
\Sigma_{\sigma_{u21}} \Sigma_{\sigma_{u11}}^{-1} u_1 &= \sigma_u^2 \left( \frac{\rho(w_{12} + w_{21})}{1 - \rho^2 w_{12} w_{21}} \right) \frac{1}{\sigma_u^2 \left( \frac{1 + \rho^2 w_{12}^2}{1 - \rho^2 w_{12} w_{21}} \right)} u_1^* = \frac{\rho(w_{12} + w_{21}) u_1^*}{1 + \rho^2 w_{21}^2} \\
\Sigma_{\sigma_{u22}} - \Sigma_{\sigma_{u21}} \Sigma_{\sigma_{u11}}^{-1} \Sigma_{\sigma_{u12}} &= \sigma_u^2 \left( \frac{1 + \rho^2 w_{21}^2}{1 - \rho^2 w_{12} w_{21}} \right)^2 \frac{\left( \frac{\rho(w_{12} + w_{21})}{1 - \rho^2 w_{12} w_{21}} \right)^2}{1 + \rho^2 w_{12}^2} = \frac{\sigma_u^2}{1 + \rho^2 w_{12}^2} \\
u_2^* | u_1^* &\sim N \left( \frac{\rho(w_{12} + w_{21}) u_1^*}{1 + \rho^2 w_{21}^2}, \frac{\sigma_u^2}{1 + \rho^2 w_{12}^2} \right) \quad (106)
\end{align*}
\]

For \( n = 3 \):

\[
\Sigma_{\sigma_u} = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{bmatrix}
\]

Let

\[
A = \begin{bmatrix}
1 & -\rho w_{12} & -\rho w_{13} \\
-\rho w_{21} & 1 & -\rho w_{23} \\
-\rho w_{31} & -\rho w_{32} & 1
\end{bmatrix} \quad (107)
\]

\[
D = \text{det}(A) = (1 - \rho^2 w_{23} w_{32}) + \rho w_{12} (-\rho w_{21} - \rho^2 w_{23} w_{31}) - \rho w_{13} (\rho^2 w_{21} w_{32} + \rho w_{31})
\]

\[
D = (1 - \rho^2 (w_{23} w_{32} + w_{12} w_{21} + w_{13} w_{31})) - \rho^3 (w_{12} w_{23} w_{31} + w_{13} w_{21} w_{32}) \quad (108)
\]
\[
\Sigma_{\sigma_w} = \sigma_n^2 [A' A]^{-1} = \begin{bmatrix}
\sigma_{a_{11}}^2 & \sigma_{a_{12}}^2 & \sigma_{a_{13}}^2 \\
\sigma_{a_{21}}^2 & \sigma_{a_{22}}^2 & \sigma_{a_{23}}^2 \\
\sigma_{a_{31}}^2 & \sigma_{a_{32}}^2 & \sigma_{a_{33}}^2
\end{bmatrix}
\] (109)

where:

\[
a_{11} = (1 - \rho^2 w_{23} w_{32})^2 + (\rho w_{12} + \rho^2 w_{13} w_{32})^2 + (\rho^2 w_{12} w_{23} + \rho w_{13})^2
\]

\[
a_{12} = (1 - \rho^2 w_{23} w_{32})(\rho w_{21} + \rho^2 w_{23} w_{32}) + (1 - \rho^2 w_{13} w_{31})(\rho w_{12} + \rho^2 w_{13} w_{32}) +
(\rho^2 w_{12} w_{23} + \rho w_{13})(\rho w_{23} + \rho^2 w_{13} w_{21})
\]

\[
a_{13} = (1 - \rho^2 w_{23} w_{32})(\rho^2 w_{21} w_{32} + \rho w_{31}) + (\rho w_{32} + \rho^2 w_{12} w_{31})(\rho w_{12} + \rho^2 w_{13} w_{32}) +
(\rho^2 w_{12} w_{23} + \rho w_{13})(1 - \rho^2 w_{12} w_{21})
\]

\[
a_{21} = (1 - \rho^2 w_{23} w_{32})(\rho w_{21} + \rho^2 w_{23} w_{32}) + (1 - \rho^2 w_{13} w_{31})(\rho w_{12} + \rho^2 w_{13} w_{32}) +
(\rho^2 w_{12} w_{23} + \rho w_{13})(\rho w_{23} + \rho^2 w_{13} w_{21})
\]

\[
a_{22} = (\rho w_{13} w_{31} + \rho^2 w_{23} w_{32})^2 + (1 - \rho^2 w_{13} w_{31})^2 + (\rho^2 w_{13} w_{21} + \rho w_{32})^2
\]

\[
a_{23} = (\rho^2 w_{23} w_{31} + \rho w_{21})(\rho w_{31} + \rho^2 w_{21} w_{32}) + (1 - \rho^2 w_{13} w_{31})(\rho w_{32} + \rho^2 w_{12} w_{31}) +
(\rho^2 w_{13} w_{21} + \rho w_{32})(1 - \rho^2 w_{12} w_{21})
\]

\[
a_{31} = (1 - \rho^2 w_{23} w_{32})(\rho^2 w_{21} w_{32} + \rho w_{31}) + (\rho w_{32} + \rho^2 w_{12} w_{31})(\rho w_{12} + \rho^2 w_{13} w_{32}) +
(\rho^2 w_{12} w_{23} + \rho w_{13})(1 - \rho^2 w_{12} w_{21})
\]

\[
a_{32} = (\rho^2 w_{23} w_{31} + \rho w_{21})(\rho w_{31} + \rho^2 w_{21} w_{32}) + (1 - \rho^2 w_{13} w_{31})(\rho w_{32} + \rho^2 w_{12} w_{31}) +
(\rho^2 w_{13} w_{21} + \rho w_{32})(1 - \rho^2 w_{12} w_{21})
\]

\[
a_{33} = (\rho w_{31} + \rho^2 w_{21} w_{32})^2 + (1 - \rho^2 w_{12} w_{21})^2 + (\rho^2 w_{12} w_{31} + \rho w_{32})^2
\]
let

\[
\sigma_{ij} = \frac{\sigma_{a_i}^2}{D^2} \text{ for } (i = 1, 2, 3 \text{ and } j = 1, 2, 3)
\]  

(110)

For \( n = 3 \), let \( \underline{u}_2 \) be a \( 2 \times 1 \) vector such that \( \underline{u}_2^* = \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} \)

\[
\underline{u}_2^* \sim N(\sigma_{a_1}^{-1} \Sigma_{\underline{u}_2}^{-1} \underline{u}_2, \sigma_{a_2}^{-1} \Sigma_{\underline{u}_2}^{-1} \Sigma_{\underline{u}_2}^{-1} \Sigma_{\underline{u}_2})
\]  

(111)

\[
\Sigma_{\underline{u}_2} = \begin{bmatrix} u_1^* & u_2^* \\ u_1^* & u_2^* \end{bmatrix} = \begin{bmatrix} u_1^* u_1' & u_1^* u_2' \\ u_2^* u_1' & u_2^* u_2' \end{bmatrix} = \begin{bmatrix} \sigma_{a_1} & \sigma_{a_1} \\ \sigma_{a_2} & \sigma_{a_2} \end{bmatrix}
\]

\[
\Sigma_{\underline{u}_2} = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1}^2 \\ \sigma_{a_2}^2 & \sigma_{a_2}^2 \end{bmatrix}
\]

\[
\Sigma_{\underline{u}_2}^{-1} = \begin{bmatrix} \frac{-a_{22}D^2}{\sigma_0^2(a_1 \alpha_2 - a_{12}^2)} & \frac{-a_{12}D^2}{\sigma_0^2(a_1 \alpha_2 - a_{12}^2)} \\ \frac{-a_{12}D^2}{\sigma_0^2(a_1 \alpha_2 - a_{12}^2)} & \frac{a_{11}D^2}{\sigma_0^2(a_1 \alpha_2 - a_{12}^2)} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}
\]
\[
\begin{align*}
\sigma_{32}^{-1} & = \begin{bmatrix} \text{cov}(u_3^*u_1^*) & \text{cov}(u_3^*u_2^*) \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} \\
\sigma_{31}^{-1} & = \begin{bmatrix} \sigma_{31} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} \\
\sigma_{32}^{-1} & = \begin{bmatrix} \sigma_{31} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} \\
\sigma_{33}^{-1} & = \sigma_{33} - \begin{bmatrix} \text{cov}(u_3^*u_1^*) & \text{cov}(u_3^*u_2^*) \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \text{cov}(u_3^*u_1^*) \\ \sigma_{21} \end{bmatrix} \\
\sigma_{33}^{-1} & = \sigma_{33} - \begin{bmatrix} \sigma_{31} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \sigma_{31} \\ \sigma_{32} \end{bmatrix} \\
\sigma_{33}^{-1} & = \sigma_{33} - (\sigma_{31}\sigma_{11} + \sigma_{32}\sigma_{21})\sigma_{31} + (\sigma_{31}\sigma_{12} + \sigma_{32}\sigma_{22})\sigma_{32} \\
\sigma_{33}^{-1} & = \frac{a_{33}^2}{D^2} - \frac{a_{31}^2}{D^2}(a_{11}a_{22} - a_{12}^2) + (a_{31}a_{12} + a_{32}a_{11})D \\
\end{align*}
\]

\[u_3^* | u_2^* \sim N\left(\frac{(a_{31}a_{22} - a_{32}a_{21})u_1^* + (a_{32}a_{12}^2 - a_{31}a_{11}a_{22} - a_{12}^2)u_2^*}{a_{11}a_{22} - a_{12}^2}, \sigma_u^2 \frac{a_{33}^2}{D^2} (a_{33}^2 - (a_{33}^2 - (a_{31}a_{12} + a_{32}a_{11})a_{32}) \right)\]

Note that when \( \rho = 0 \Rightarrow \sigma_{ij} = 0 \) for \( i \neq j \) and \( \sigma_{ij} = 1 \) for \( i = j \), and \( D = 1 \). For \( u_2^* | u_1^* \)
\[ u_2^* | u_1^* \sim \mathcal{N}(\sigma_{21} \Sigma_{\sigma_{u11}}^{-1} u_1, \sigma_{22} - \sigma_{21} \Sigma_{\sigma_{u11}}^{-1} \sigma_{12}) \] (113)

\[
\sigma_{21} = \frac{a_{21} \sigma_u^2}{D^2}, \quad \Sigma_{\sigma_{u11}}^{-1} = \frac{1}{a_{11} \sigma_u^2}
\]

\[
\Rightarrow \sigma_{21} \Sigma_{\sigma_{u11}}^{-1} u_1 = \frac{a_{21} \sigma_u^2}{D^2} u_1 = \frac{a_{21} u_1}{a_{11}}
\]

\[
\sigma_{22} - \sigma_{21} \Sigma_{\sigma_{u11}}^{-1} \sigma_{12} = \frac{a_{22} \sigma_u^2}{D^2} - \frac{a_{11} a_{21} \sigma_{12}}{D^2} = \frac{a_{22} a_{11} - a_{12}^2}{D^2 a_{11}}
\]

\[
u
\sigma_{22} - \sigma_{21} \Sigma_{\sigma_{u11}}^{-1} \sigma_{12} = \frac{a_{22} a_{11} - a_{12}^2}{D^2 a_{11}}
\]

\[ u_2 | u_1 \sim \mathcal{N}\left(\frac{a_{21} u_1}{a_{11}}, \frac{\sigma_u^2 (a_{22} a_{11} - a_{12}^2)}{D^2 a_{11}}\right) \] (114)
### Table 1

<table>
<thead>
<tr>
<th>( MLEs(Avg) ) ( \rho = -0.1 )</th>
<th>( \rho = 0.2 )</th>
<th>( \rho = 0.4 )</th>
<th>( \rho = 0.6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_u = 0.1964 ) ( \hat{\beta} )</td>
<td>0.9978</td>
<td>1.5850</td>
<td>1.9939</td>
</tr>
<tr>
<td>( \hat{\sigma}_v )</td>
<td>0.5882</td>
<td>0.4264</td>
<td>0.5968</td>
</tr>
<tr>
<td>( \hat{\sigma}_u )</td>
<td>0.3926</td>
<td>0.3281</td>
<td>0.5739</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>-0.4150</td>
<td>-0.4332</td>
<td>0.0457</td>
</tr>
</tbody>
</table>

\( \sigma_u = 0.4472 \)

| \( \hat{\beta} \) | 0.7102 | 1.1372 | 0.7553 | -0.2924 |
| \( \hat{\sigma}_v \) | 0.5058 | 0.5498 | 0.0001 | 0.3803 |
| \( \hat{\sigma}_u \) | 0.3318 | 0.3086 | 0.7364 | 0.7389 |
| \( \hat{\rho} \) | -0.6813 | -0.6467 | -1.0337 | -0.2453 |

\( \sigma_u = 0.7071 \)

| \( \hat{\beta} \) | 1.8383 | 0.6981 | 0.4078 | -1.1637 |
| \( \hat{\sigma}_v \) | 0.3087 | 0.5115 | 0.6591 | 0.3734 |
| \( \hat{\sigma}_u \) | 0.1656 | 0.1111 | 0.3102 | 1.0368 |
| \( \hat{\rho} \) | -0.8896 | 0.1167 | 0.0907 | 0.2927 |

\( \sigma_u = 0.8944 \)

| \( \hat{\beta} \) | 1.2301 | 1.0020 | 0.0745 | -0.7155 |
| \( \hat{\sigma}_v \) | 0.3921 | 0.2927 | 0.0514 | 0.1512 |
| \( \hat{\sigma}_u \) | 0.1342 | 0.1405 | 0.3971 | 0.6718 |
| \( \hat{\rho} \) | -0.5286 | 0.3369 | -0.6582 | -0.1359 |

**END OF APPENDIX**
Vita

Syracuse University
Department of Economics
110 Eggers Hall
Syracuse, NY 13244

Mobile: (914)803-9852
Email: iawright@syr.edu
Website: https://iawright.expressions.syr.edu/
Citizenship: United States of America

EDUCATION:
M.A., Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA May 2006
M.Sc., Economics, The University of the West Indies, Kingston, Jamaica May 2002
B.Sc., Double Major, Economics and Statistics, The University of the West Indies, Kingston, Jamaica First Class Honors (summa cum laude) May 2000

PUBLICATION (S):

Peer-reviewed manuscripts:

PRESENTATION (S):


WORK EXPERIENCE:

Consultant

Economic, Planning and Research Division
Economist

Planning Institute of Jamaica May 2001 – Aug. 2001
Economic, Planning and Research Division
Research Assistant

University of Massachusetts Boston, Boston, MA July 2014 – May 2015
Economics Department
Part-time Faculty
Syracuse University, Syracuse, NY  
Graduate Assistant  
Spring 2014; Summer 2013; Fall 2013  
Summer 2012; Aug. 2011–May 2012; Fall 2010 – Spring 2011

Virginia Polytechnic Institute and State University, Blacksburg, VA  
Teaching Assistant  
Fall 2004–Spring 2006

University of the West Indies, Kingston, Jamaica  
Teaching Assistant  
Fall 2002  
Fall 2000–Spring 2001

FELLOWSHIP AND AWARDS:

Ford Foundation Dissertation Fellowship, Alternate, Honorable Mention  

Science, Technology, Engineering and Math (STEM) Fellowship,  
Syracuse University  
Aug. 2012–May 2013  
Aug. 2009–May 2010

Maxwell School Summer Fellowship, Syracuse University  
Summer 2012

Summer Research Grant, Virginia Polytechnic Institute and State University  
Summer 2006

Starr Foundation Fellowship, The University of the West Indies  
Aug. 2001–May 2002

Certificate of Participation, Introduction to teaching skills workshop,  
The University of the West Indies  
May 2000

Certificate of Merit, Quality Leadership Program,  
The University of the West Indies  
May 1999

West Indies Guild of University Teachers (WIGUT) Academic Bursary,  
The University of the West Indies  

Bank of Novia Scotia Academic Bursary,  
The University of the West Indies  

SKILLS:

Statistical Software Packages: EVIEWS, STATA, MatLab, Scientific Workplace, SPSS and Microsoft Office XP