Production Analysis with Asymmetric Noise

Badunenko, Oleg and Henderson, Daniel J.

Brunel University London, University of Alabama

31 November 2021

Production Analysis with Asymmetric Noise*

Oleg Badunenko[⋆] Daniel J. Henderson[§]

November 2021

Abstract

Symmetric noise is the prevailing assumption in production analysis, but it is often violated in practice. Not only does asymmetric noise cause least-squares models to be inefficient, it can hide important features of the data which may be useful to the firm/policymaker. Here we outline how to introduce asymmetric noise into a production or cost framework as well as develop a model to introduce inefficiency into said models. We derive closed-form solutions for the convolution of the noise and inefficiency distributions, the log-likelihood function, and inefficiency, as well as show how to introduce determinants of heteroskedasticity, efficiency and skewness to allow for heterogenous results. We perform a Monte Carlo study and profile analysis to examine the finite sample performance of the proposed estimators. We outline R and Stata packages that we have developed and apply to three empirical applications to show how our methods lead to improved fit, explain features of the data hidden by assuming symmetry, and how our approach is still able to estimate efficiency scores when the least-squares model exhibits the well-known "wrong skewness" problem in production analysis.

Keywords: asymmetry, production, cost, efficiency, wrong skewness
JEL: C13, C21, D24, I21

---

* This paper benefited from the feedback provided by colleagues and participants of conferences. We specifically wish to thank Taras Bodnar, William Horrace, Subal Kumbhakar, Pavlo Mozharovskyi, Alecos Papadopoulos, Christopher Parmeter, and Léopold Simar. We wish to thank Seth Gershenson for providing data.

⋆ Department of Economics and Finance, College of Business, Arts and Social Sciences, Brunel University London, Uxbridge, UB8 3PH, UK; oleg.badunenko@brunel.ac.uk

§ Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, 35487-0224, USA; djhender@cba.ua.edu

1 Introduction

The usage of symmetric noise in econometric models engraves the assumption that a circumstance, choice, or behavior of an economic agent is ruled by the normal law where positive and negative deviations from the "trend" have the same effect on the outcome variable of an individual firm. There is substantial evidence across a wide array of fields to suggest that in practice, symmetry is not always a reasonable assumption (e.g., Genton, 2004).

Asset Pricing The probability distribution of asset returns is often skewed (Adcock, 2007). When the distribution is asymmetric, the mean and variance are not sufficient statistics for investors to make optimal asset allocation decisions and ordinary least-squares (OLS) estimation is inefficient. Hence, authors have looked to methods that exploit the asymmetric nature of the data. For example, Adcock (2005, 2010) employ multivariate skewed distributions to study the sensitivity of asset returns to return on the market portfolio. These methods extend the mean-variance methods for portfolio selection to mean-variance-skewness which can lead to improvements in performance.

Risk Management A popular measure of an investment prospect is Value at Risk (VaR), which measures the risk of loss for investments. It is obtained by focusing on the the bottom tail of the returns distribution. For simplicity and convenience, it is often naively assumed that the distribution is symmetric. Assuming symmetry can vastly understestimate the risk being taken on by the investor. Exploiting the asymmetric nature of the data can lead to gains. For example, Goh et al. (2012) are able to outperform mean-variance approaches using half-space statistical information when asset returns are asymmetric.

Banking Asymmetric shocks can severely impact banks. For example, managers may take on excess risk as a consequence of a principal agent problem. These low probability events emerge as large negative shocks. On the other side, deposits across large banks and savings institutions or within a single bank are highly positively skewed (Aubuchon and Wheelock, 2010). In practice, the direction of the asymmetry may not be clear *a priori*.

Supply Shocks Ball and Mankiw (1995) study the effects of supply shocks on inflation (i.e., shifts in the short-run Philips curve) based on relative price changes and frictions in nominal price adjustments. Price rigidities typically occur because of a sluggish price adjustment and costs associated with adjusting nominal prices. Firms typically adjust to large shocks, but not to small shocks and thus these large shocks have a disproportional impact on prices. The authors argue in favor of disproportionate effects of supply shocks on inflation and find that the inflation-skewness relationship is stronger than the inflation-variance relationship.

Interest Rate Parity Louis et al. (1999) account for transaction costs in testing interest rate parity (IRP). They consider the relevant no-arbitrage conditions that in equilibrium are bounded in one direction. They argue that the assumption of symmetric noise in an IRP equation would result in inconsistency and therefore consider skewed composite errors (convolution of symmetric noise and one-sided error term) using stochastic frontier analysis. With this approach, they find that arbitrage margins are sometimes violated and hence there are possible arbitrage opportunities.

Educational Outcomes There is a large literature estimating production functions using educational data (e.g., Ruggiero, 1996; Thanssoulis, 1999; de Witte et al., 2010; Thanassoulis et al., 2016; Johnes et al., 2017; de Witte and López-Torres, 2017; Thanassoulis et al., 2017, 2018). Random shocks occur in education and some of these can be unpleasant or terrible events such as bullying, bereavement, unfair treatment, or an external event such as a school shooting. In practice, it is common to model these "shocks" as inputs in an educational production function (Ponzo, 2013). Alternatively, we can allow for these large negative impacts on educational outcomes (Gershenson and Tekin, 2018) to be treated as shocks. We no longer need to attribute such observations as outliers, because asymmetric noise distributions can potentially account for these unfortunate events.

Weather Asymmetry in weather shocks also plays a role in production. Floodings, droughts, tornadoes and earthquakes are thought of as low probability events, but can result in huge damages. In agriculture, adverse events play an important role as they jeopardize the harvest. For example, Qi et al. (2015) use climactic variables as inputs in a stochastic production frontier of Wisconsin dairy farms. Modeling these events via asymmetric shocks may help determine potential losses which may prove useful for crop insurance premiums (Shaik, 2013).

1.1 Modeling Asymmetry in Production

If such asymmetries in preferences and behavior are not accounted for, we may estimate the wrong model that eventually leads to incorrect policy prescriptions. For example, the normality of the crop yield has long been rejected and was shown to be skewed, and ignoring this lead to overprediction of field crop yields (Day, 1965). Profits can be driven by asymmetric capacities (Mao et al., 2019).

The main goal of this paper is to model production uncertainty by allowing for asymmetric noise in production analysis. Asymmetric noise is extremely rare in production economics.[1] We propose a set of models that introduce asymmetric noise in estimation of a production relationship in situations where a researcher believes that the production units may operate with or without efficiency.

1.2 Inefficiency

In empirical applications, firms or individuals are often assumed to operate with 100% efficiency. In neoclassical economics, firms and economic agents can exhibit inefficiency by being below their production possibilities. The conceptualization and formulation of inefficiency in production can be traced back to Koopmans (1951) and Afriat (1972).

In their seminal econometric papers, Aigner et al. (1977) and Meeusen and van den Broeck (1977) formulated stochastic frontier (SF) models, where inefficiency followed a half-normal

[1] Bonanno et al. (2017) consider a generalized logistic distribution for noise and Wei et al. (2021) (independently) consider a skew-normal distribution. The set of models that we introduce in this article go a step further by deriving closed-form solutions and introducing determinants into each of the components. Models that introduce technical inefficiency into a production process (Domínguez-Molina et al., 2004), which we discuss next, are also exceptions. However, as we show below, the latter models are special cases of a model with a general asymmetric noise.

and exponential distribution, respectively. Many extensions of these models exist and include other distributions for the unobserved inefficiency component. These include assuming the distribution of inefficiency to be truncated normal (Stevenson, 1980), truncated normal with determinants (Kumbhakar et al., 1991), jointly estimated technical and allocative efficiency (Kumbhakar and Tsionas, 2005), generalized exponential distribution of inefficiency (Papadopoulos, 2021), semiparametric smooth coefficient framework (Yao et al., 2019), dealing with endogeneity (Amsler et al., 2016; Lien et al., 2018; Lai and Kumbhakar, 2018) and modeled where noise can follow any (symmetric) law (Florens et al., 2020). Greene (2008) and Stead et al. (2019) discuss methodological advances in stochastic frontier modeling and especially distributional specifications. While in academic papers the convolution of the noise and inefficiency distributions are overwhelming skewed (Li, 1996), each of these models assumes that the noise is term is symmetrically (overwhelmingly normally) distributed (Horrace and Parmeter, 2018; Wheat et al., 2019).

Here we will propose a SF model whereby the skewness of the *composite* error (convolution of noise and inefficiency) may have either sign. We formulate a composite error that is skew-normal for the noise and has a one-sided distribution for the inefficiency component. We are able to derive closed-form solutions for the convolution of the two distributions as well as the log-likelihood function and its gradients. Further, we derive closed-form solutions for the inefficiency estimates as well as discuss how to incorporate determinants of heteroskedasticity, efficiency and skewness to allow for heterogenous effects.

It turns out, with our approach, if we take the stance that "wrong skewness" is an empirical issue (e.g., Simar and Wilson, 2009), we are still able to estimate efficiency scores when least-squares residuals are of the "wrong skewness" (Olson et al., 1980; Cho and Schmidt, 2020). "Wrong skewness" is an empirical artifact that occurs when least-squares residuals have a positive skew in a production function or negative skew in a cost function. In this case, the SF model is inconsistent with the data and it is assumed that there is no inefficiency.

1.3 Finite Sample Performance

Obviously, all of our parameters are identified by the parametric assumptions on the model and the maximum likelihood principle, however, in practice, it can sometimes be difficult to estimate parameters via standard maximum likelihood techniques. This seems especially true when we have two forms of asymmetry and the sign of one of those is potentially unknown. To understand how our estimators perform in various scenarios and with various sample sizes, we conduct a Monte Carlo study and profile analysis. We obtain reliable estimates of the variance parameters in all scenarios and reliable estimates of our skewness parameter for sample sizes at or above above 200. In short, our study suggests that our estimators possess desirable finite sample properties.

1.4 Empirical Performance

In order to see how asymmetric noise distributions perform in practice, we provide three empirical applications. The first application looks at risk behavior of U.S. Banks. We re-examine the cost function in Restrepo-Tobón and Kumbhakar (2014) both with and without assuming symmetric noise. Our metrics suggest that the SF model with skewed noise best fits the data. We further discover that the most risky banks (as determined by the standard deviation of return on assets) are more likely to be hit by negative shocks that have large negative effects on total costs.

In our second example, we look at an educational production function. Here we take the data collected by Gershenson and Tekin (2018) to see the impact the "Beltway Sniper" had on public school student math test scores in Virginia. As in the previous example, our SF model with asymmetric noise best fits the data. Here we find that skewness of the noise distribution is negative and is getting closer to 0 as a school is further away from a sniper attack. In other words, for those schools that are close to at least one sniper attack scene, they have a larger probability to exhibit poor academic performance.

In these first two examples, we demonstrate the performance of the proposed models applied to cost and production functions, respectively. We conclude that the model that takes the skewness of the noise distribution into account is superior to the model with symmetric noise. In both applications, we find that the most flexible model that allows (i) skewed noise where (ii) the parameters of its distribution vary across observations as well as (iii) inefficiency with observation-specific determinants performs best and provides the richest scope for interpretation. We expect that these methods will prove fruitful in uncovering previously ignored/misplaced information.

In our final example, we take data from the NBER-CES Manufacturing Industry Database (Bartelsman and Gray, 1996) and examine the efficiency scores of 4-digit textile industries. For each year (1958–2011), we run separate cross-sectional regressions and report both the estimated skewness parameter and average efficiency score for each year. While most skewness estimates are near zero, many estimates are significantly above or below zero. In those years where the model has the "wrong skewness", the conventional SF model predicts no inefficiency. This is found to be the case in about half of the cases. For our SF model, for those years, the average estimated efficiency scores are below unity.

1.5 Roadmap

The remainder of the paper is organized as follows: Section 2 summarizes the skew-normal distribution. Section 3 proposes to allow for skew-normal noise in a production or cost function as well as extends the model to allow for inefficiency. This section further examines the finite sample performance of our estimators and how to implement the procedures in both `R` and Stata with packages that we have created. Section 4 provides our empirical examples and the fifth section concludes. The appendices include our full set of derivations (Appendix A), extensions to truncated normal inefficiency (Appendix B), the results of the simulation study and profiling analysis (Appendix C) as well as `R` code to help replicate our empirical and simulation results (Appendix D).

2 Skew-Normal Distribution

In what follows, we employ a skew-normal (SN) noise distribution. While other distributions may be feasible or more general, we chose this skewed distribution for at least five reasons. First, it is a well studied skewed distribution with known properties and inferential aspects. Second, the standard model with normally distributed noise is a special case of the SN. Third, we are able to derive closed form solutions for many objects of interest. Fourth, it can be skewed in either direction and only requires one additional parameter to estimate.[2] Finally, our analysis can be the basis for extensions to more complicated skew-elliptical distributions (Genton, 2004; Azzalini and Capitanio, 2013).

Formally, the SN distribution generalizes the normal distribution by allowing for non-zero skewness. The probability density function of the extended SN distribution with the skewness parameters α_0 and α_1, the location parameter $\xi \in \mathbb{R}$, and the variance $\sigma_\omega^2 > 0$ is given by

$$g(\omega; \xi, \sigma_\omega^2, \alpha_0, \alpha_1) = \frac{\phi\left(\frac{\omega - \xi}{\sigma_\omega}\right) \Phi\left(\alpha_0 + \alpha_1 \left(\frac{\omega - \xi}{\sigma_\omega}\right)\right)}{\Phi\left(\frac{\alpha_0}{\sqrt{1 + \alpha_1^2}}\right)},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of a standard normal distribution, respectively. We say that ω is skew-normally distributed: $\omega \sim SN(\xi, \sigma^2, \alpha_0, \alpha_1)$.

Azzalini (1985) proposes to set $\alpha_0 = 0$ so the skewness is determined by a single parameter $(\alpha \equiv \alpha_1)$.[3] The density becomes

$$h(\omega; \xi, \sigma_\omega^2, \alpha) = \frac{2}{\sigma_\omega} \phi\left(\frac{\omega - \xi}{\sigma_\omega}\right) \Phi\left(\alpha \frac{\omega - \xi}{\sigma_\omega}\right), \tag{1}$$

[2] This convenience and simplicity comes at a price as the "SN family does not provide an adequate stochastic model for cases with high skewness or kurtosis" (Azzalini and Capitanio, 2013). That being said, the SN distribution offers a statistical model that regulates the skewness, is tractable (closed form solutions) and is easily interpretable.

[3] See Azzalini and Capitanio (2013, Chapter 2) for details on the extended skew-normal distribution.

where the expected value of $\omega \sim SN(\xi, \sigma^2, \alpha)$ is

$$E(\omega) = \xi + \sigma_\omega \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

For the case where $E(\omega) = 0$, the density in (1) can be concentrated in terms of ξ and can be written as

$$h(\omega; \xi = 0, \sigma_\omega^2, \alpha) = \frac{2}{\sigma_\omega} \phi(\omega_{rs}) \Phi(\alpha \omega_{rs}), \tag{2}$$

where the rescaled and shifted ω is given by

$$\omega_{rs} = \frac{\omega}{\sigma_\omega} + \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

The shape of the density is determined by the parameter α. The upper and lower panels of Figure 1 show densities of a SN random variable for $\sigma_\omega = 0.1$ and $\sigma_\omega = 5$. The two plots differ only by the scale of the axes. Here we choose only to show negative values of the skewness parameter ($\alpha < 0$). For positive values of α, the density is flipped symmetrically around 0. As the absolute value of α increases, the skewness of the distribution is increasing. For $\alpha = \infty$, the skew normal distribution becomes the truncated normal (Horrace, 2005a,b). Figure 1 suggests that the distribution is very skewed (i.e., approaches the truncated normal distribution) for an absolute value of α around 10.[4]

3 Production Model

In this section, we describe how to introduce an asymmetric noise distribution into a production framework. We then derive the results for this noise distribution in a stochastic frontier framework. More specifically, we derive closed form solutions for the convolution of the noise and inefficiency distributions, the log-likelihood function, and inefficiency, as well show how to introduce determinants of heteroskedasticity, efficiency and skewness to allow for heterogenous

[4] See DiCiccio and Monti (2004) for inferential aspects of the parameters of the SN distribution.

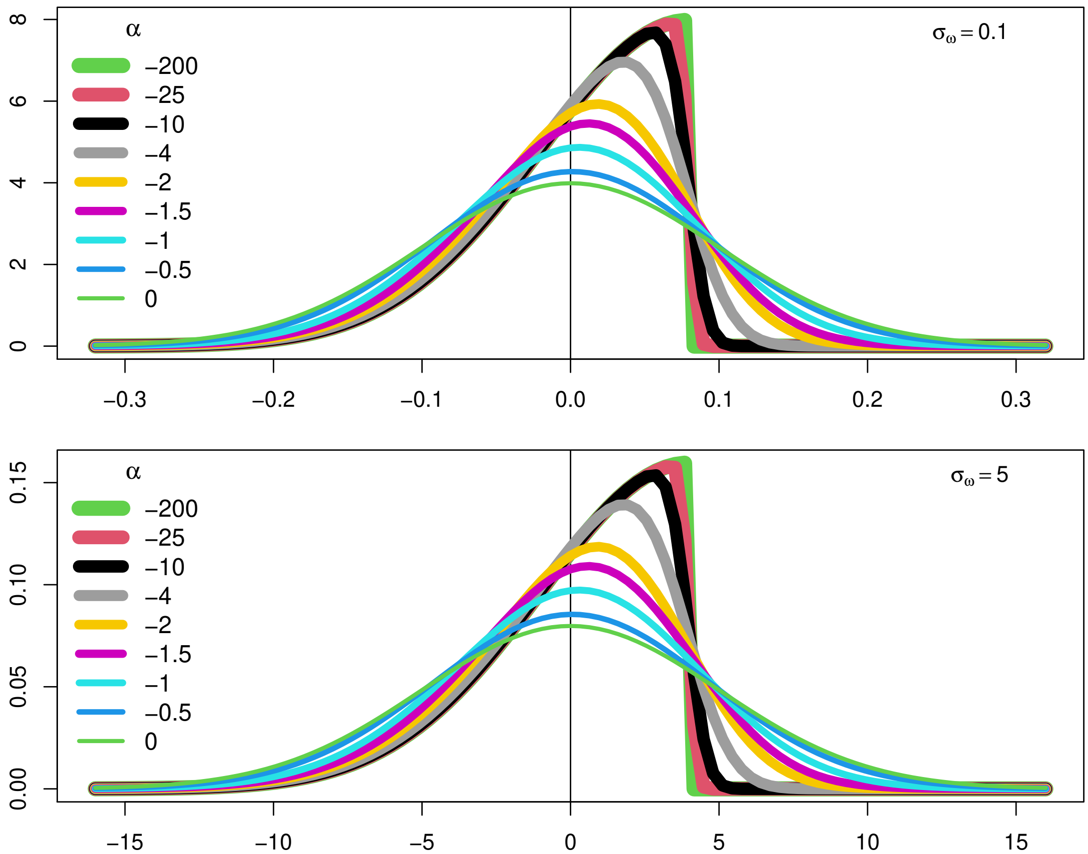


Figure 1: pdf of the Skew-Normal Random Variable

results. Finally, we discuss finite sample performance via a Monte Carlo and profile analysis as well as mention R and Stata packages that we have developed and will distribute so that our results may be replicated and for authors to use for their own studies.

Our production function can be written as

$$y = f(\boldsymbol{x}; \boldsymbol{\beta}) + v, \tag{3}$$

where the outcome variable y is the logarithm of output for a stochastic production function (or the logarithm of cost for a stochastic cost function). $f(\boldsymbol{x}; \boldsymbol{\beta})$ is a log-linear (in parameters) production or cost function with input row vector $\boldsymbol{x}$ (a constant, logarithms of the input variables and possibly other observed covariates that include environmental variables that are not primary inputs, but nonetheless affect the outcome variable) and the finite parameter vector $\boldsymbol{\beta}$ (Sun et al., 2011).

We assume that the noise v is SN distributed with zero expectation, $E(v) = 0$,

$$v \sim SN\left(-\sigma_v\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}, \sigma_v^2, \alpha\right)$$

with a probability density function (pdf) adopted from equation (2)

$$f_v(v) = \frac{2}{\sigma_v}\phi\left(\frac{v}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right)\Phi\left(\alpha\left[\frac{v}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right]\right).$$

The log-likelihood function for a log-linear (in parameters) conditional expectation production (or cost) function with SN noise is given as

$$\begin{aligned} \ln 2 - \ln\sigma_v + \ln\left[\phi\left(\frac{y - f(\boldsymbol{x};\boldsymbol{\beta})}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right)\right] \\ + \ln\left[\Phi\left(\alpha\frac{y - f(\boldsymbol{x};\boldsymbol{\beta})}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha^2}{\sqrt{1+\alpha^2}}\right)\right] \end{aligned} \tag{4}$$

and the parameters can be estimated via maximum-likelihood (ML). Note that while least-squares estimation here is unbiased as it is equivalent to the quasi-maximum likelihood estimator under the assumption of normally distributed errors, it is no longer efficient (Yao and Zhao, 2013).

Here we note the relationship of what we have just presented to the model originally proposed by Aigner et al. (1977), where the error term v in (3) is composed of a symmetric component that is normally distributed with a variance ς^2 and a non-negative technical inefficiency component that is half-normally distributed with variance τ^2. Replacing α in (4) by τ/ς and σ_v by $\sqrt{\tau^2+\varsigma^2}$ yields the likelihood function for the model proposed by Aigner et al. (1977) (see equation (13.2) in Domínguez-Molina et al. (2004) as well as the discussion in Badunenko and Kumbhakar (2016)). In other words, the popular SF model can be seen as a special case of the model considered in (3). The inferential aspects of this special case were studied in Badunenko et al. (2012).

With the exception of Li (1996), SF models employ an asymmetric compound noise. However, those models make an assumption that the asymmetry that is present in the composite error term is due to existing technical inefficiencies. We propose a set of models where we split the asymmetry/skewness into components attributable to uncertainty (skewed noise) and technical inefficiency (non-negative error part). We show that they can be separated. In what follows, we present a more general model, where inefficiency exists and the noise can be skewed.

3.1 Production Model with Inefficiency

In the presence of inefficiency, (3) becomes

$$y = f(\boldsymbol{x};\boldsymbol{\beta}) + v - \mathsf{p}u = f(\boldsymbol{x};\boldsymbol{\beta}) + \epsilon, \tag{5}$$

where, analogous to before, the outcome variable y is the logarithm of output for a stochastic production frontier model or the logarithm of cost for a stochastic cost frontier model, $\boldsymbol{x}$ is the row vector of a constant, logarithms of the input variables and possibly other observed covariates that include environmental variables that are not primary inputs but nonetheless affect the outcome variable. To present this in a general setting, we introduce the known value p, which signifies either a production or cost function:

$$\mathsf{p} = \begin{cases} 1 & \text{for a stochastic production frontier model} \\ -1 & \text{for a stochastic cost frontier model.} \end{cases}$$

We assume that the noise v is SN distributed with a zero expectation, $E(v) = 0$,

$$v \sim SN\left(-\sigma_v\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}, \sigma_v^2, \alpha\right)$$

with a pdf adopted from equation (2)

$$f_v(v) = \frac{2}{\sigma_v}\phi\left(\frac{v}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right)\Phi\left(\alpha\left[\frac{v}{\sigma_v} + \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}\right]\right).$$

We assume that the inefficiency term is exponentially distributed (Jradi et al., 2021), so its density is given by

$$f_u(u) = \lambda \exp(-\lambda u),$$

where $\lambda = \frac{1}{\sigma_u}$.[5] Denoting $\xi_v = -\sigma_v\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}$ and noting from equation (5) that $\epsilon = v - \mathsf{p}u$, and $v - \xi_v = \epsilon + pu - \xi_v = \epsilon_r + \mathsf{p}u$, where $\epsilon_r = \epsilon - \xi_v$, the joint density of u and ϵ is given by

$$\begin{aligned} f(\epsilon, u) &= \underbrace{\frac{2}{\sigma_v}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2\right]\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)}_{f_v(v)}\underbrace{\lambda\exp(-\lambda u)}_{f_u(u)} \\ &= \frac{2\lambda}{\sigma_v}\frac{1}{\sqrt{2\pi}}\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)\exp\left[-\frac{1}{2}\left\{\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2 + 2\lambda u\right\}\right]. \end{aligned} \tag{6}$$

3.1.1 Convolution of the Skew Normal and Exponential Distributions

The marginal density of ϵ is obtained by integrating u out of $f(\epsilon, u)$, noting that $u \geq 0$ (i.e., $f(\epsilon) = \int_0^\infty f(\epsilon, u)du$). To do so, we first rewrite equation (6) as

$$f(\epsilon, u) = \frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right).$$

Then,

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(\epsilon, u)du \\ &= \int_0^\infty \frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)du \\ &= \frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\int_0^\infty \phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)du. \end{aligned}$$

5 We also considered the case of truncated normally distributed inefficiency (u) and these results are provided in Appendix B.

The integral

$$\int_0^\infty \phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)du \tag{7}$$

can be obtained in a closed form using Owen's T-function (see Owen, 1956, 1980). The details of the derivation are given in Appendix A. Denote the solution to (7) as $\mathcal{A}$:

$$\begin{aligned}\mathcal{A} = &-T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right)\\ &+ T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right)\Phi\left(-u_1\right),\end{aligned} \tag{8}$$

where $a = -\alpha\mathsf{p}\lambda\sigma_v$, $b = \alpha\mathsf{p}$, $a_2 = a/\sqrt{1+b^2}$, $u_1 = \mathsf{p}\epsilon_r/\sigma_v + \lambda\sigma_v$ and

$$T(h,a) = \frac{1}{2\pi}\int_0^a \frac{\exp\left[-0.5h^2\left(1+t^2\right)\right]}{1+t^2}dt.$$

Then the marginal density can be given in closed form as

$$f(\epsilon) = 2\lambda\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\mathcal{A}, \tag{9}$$

where examples of this probability density function for a few choices of the three parameters σ_v, α and σ_u are shown in Figure 2.

Given the above information, the log-likelihood based on (9) is

$$\ln(2\lambda) + \mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2} + \ln\mathcal{A}. \tag{10}$$

The full derivation, as well as the gradients of this log-likelihood function, which are useful for programming purposes, can be found in Appendix A.

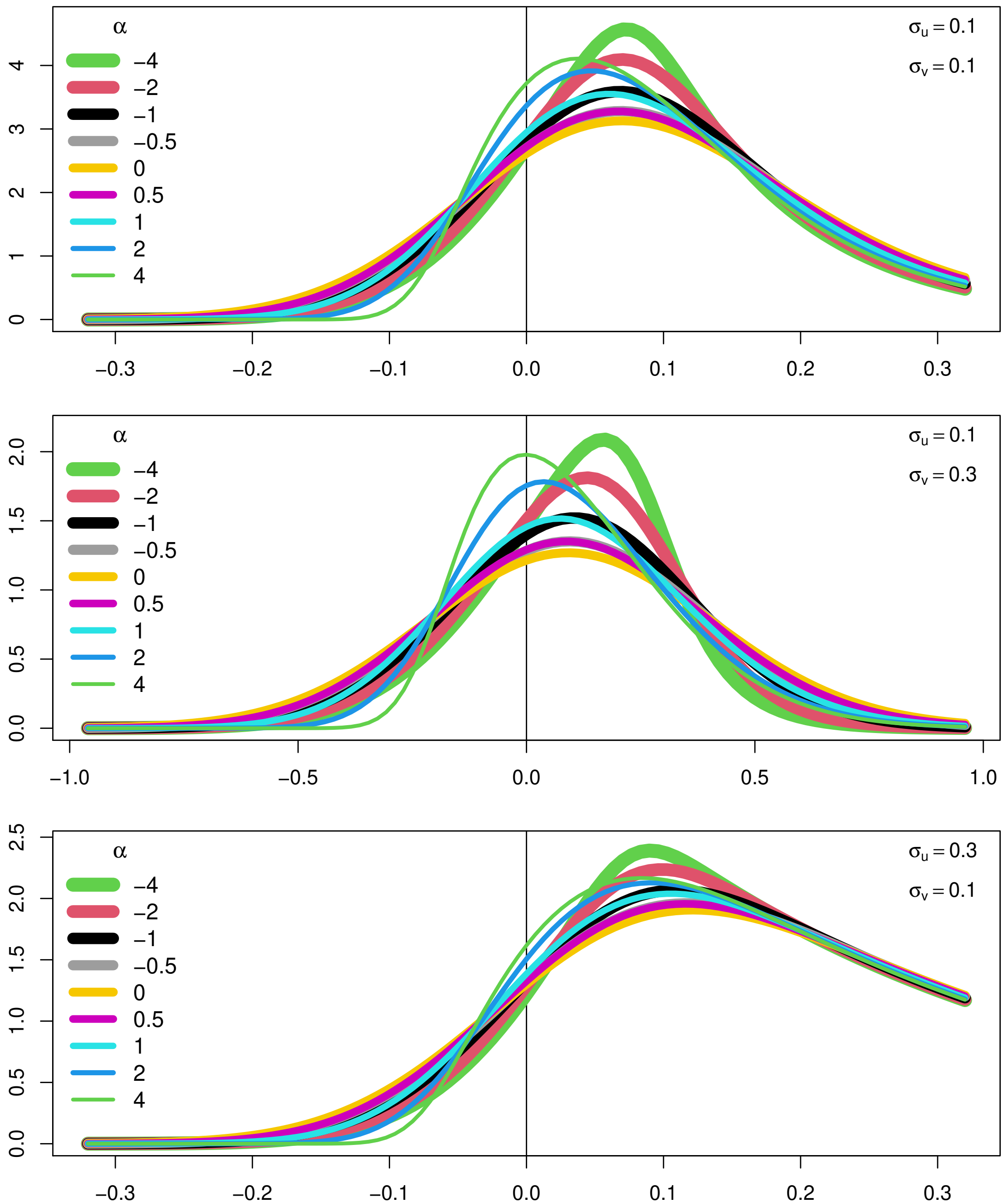


Figure 2: pdf of the Convolution of Skew-Normal and Exponential Random Variables, $\epsilon = v - \mathsf{p}u$, where $\mathsf{p} = -1$

3.1.2 Efficiency Estimation

To obtain observation-specific estimates of inefficiency (u), we follow Jondrow et al. (1982) and first obtain the conditional distribution of u given ϵ:

$$\begin{aligned} f(u|\epsilon) &= \frac{f(u,\epsilon)}{f(\epsilon)} \\ &= \frac{\frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{2\lambda\exp\left(\mathsf{p}\epsilon_r\lambda+\frac{\lambda^2\sigma_v^2}{2}\right)\mathcal{A}} \\ &= \frac{\frac{1}{\sigma_v}\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{\mathcal{A}}. \end{aligned} \quad (11)$$

We then obtain the point estimator for u (observation-specific) by finding the mean value of the conditional distribution in (11),

$$E(u|\epsilon) = \int_0^{+\infty} \frac{u \times \frac{1}{\sigma_v}\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{\mathcal{A}}du. \quad (12)$$

It can be shown (see Appendix A) that the integral in (12) has a closed form solution,

$$E(u|\epsilon) = -\mathsf{p}\epsilon_r - \lambda\sigma_v^2 + \frac{\sigma_v}{\mathcal{A}} \times \left[\begin{array}{l} \frac{b}{\sqrt{1+b^2}}\phi\left(\frac{a}{\sqrt{1+b^2}}\right) \\ \times\Phi\left(-u_1\sqrt{1+b^2} - \frac{ab}{\sqrt{1+b^2}}\right) \\ +\phi\left(u_1\right)\Phi\left(a+bu_1\right) \end{array}\right], \quad (13)$$

where $\mathcal{A}$ is defined in (8) and a, b, and u_1 are defined immediately after. The estimates of efficiency can be obtained by exponentiating the negation of the quantity in (13).

3.1.3 Determinants of Heteroskedasticity, Efficiency, and Skewness

It is feasible to modify our approach to allow for determinants of all parameters of error components (Kumbhakar et al., 1991; Lien et al., 2018). In other words, assuming data are available, we can model each component (variance, inefficiency and skewness) with both a deterministic and a stochastic component. We can attempt to explain the performance of firms based on exogenous variables within the firm's production environment.[6] Examples of naturally occurring environmental variables include, but are not limited to, human capital levels of managers, input and output quality measures, market share and/or climactic variables.

The noise term can be made heteroskedastic by allowing the variance to depend upon a set of exogenous environmental variables ($\boldsymbol{z}_v$). To ensure that the variance is positive, we adopt the following specification

$$\ln \sigma_v^2 = \boldsymbol{z}_v \boldsymbol{\gamma}_v, \tag{14}$$

where the parameter vector $\boldsymbol{\gamma}_v$ may include an intercept term. Since noise in a production relationship can be viewed as production risk, the typically employed determinant of noise variance is the size of the unit of observation (e.g., total assets in banking).

Similarly, the variance of inefficiency, and hence the inefficiency itself, can be modeled to depend upon a set of exogenous environmental variables ($\boldsymbol{z}_u$). Again, to ensure that the variance is positive, we adopt the specification

$$\ln \sigma_u^2 = \boldsymbol{z}_u \boldsymbol{\gamma}_u, \tag{15}$$

where the parameter vector $\boldsymbol{\gamma}_u$ may include an intercept term.

The first two approaches exist in the literature (Caudill et al., 1995), and here we suggest they analogously be extended for the skewness parameter to allow for heterogenous effects. Our skewness parameter can be made observation specific via

$$\alpha = \boldsymbol{z}_s \boldsymbol{\gamma}_s, \tag{16}$$

where again, the parameter vector $\boldsymbol{\gamma}_s$ may include an intercept term. Allowing for hetero-

[6] It is important to note that we will estimate all of these parameters, as well as those in the production or cost function, jointly (Wang and Schmidt, 2002; Schmidt, 2011).

geneity in skewness may be particularly useful as we may be able to determine that some firms are more susceptible to negative shocks than others. Note that this formulation allows for the skewness to take either sign and heterogeneity (as we will see later in our empirical applications) allows for both signs within a given dataset.

3.2 Finite Sample Performance

The parameters of (3) and (5) are obtained using maximum likelihood estimation (MLE) based on (4) and (10), respectively. The theoretical properties of MLE are well-known and all our parameters are identified by the parametric assumptions on the model. However, it can sometimes be difficult to obtain reliable estimates for some datasets in practice. The finite sample properties of the MLE estimator for (4) for different parameter constellations has been studied by Azzalini and Capitanio (1999) and Badunenko et al. (2012). If one considers (5) to be a generic statistical model with two skewed distributions, Badunenko and Kumbhakar (2016) studied the finite sample properties of a special case of this model.

For completeness, we have performed a small Monte Carlo study and profile analysis (Ritter and Bates, 1996). Tables with estimated bias and MSE as well as likelihood profiles are available in Appendix C. The plots of the medians of the likelihood ratio statistics show the effect the sample size has on the finite sample performance of the estimator. As expected, the parameters are more precisely estimated with larger samples. We find some evidence that α may be difficult to estimate the parameters precisely for sample sizes below 200. Further, some profiles suggest the possibility of local maxima for α (Azzalini and Capitanio, 2013, Chapter 3). To avoid this issue in practice, we suggest using a multistart procedure for optimization when using Broyden-Fletcher-Goldfarb-Shanno (BFGS) or Newton-Raphson (NR) methods.[7] The variance parameters in both (3) and (5) are precisely estimated in all scenarios. Overall, our Monte Carlo study suggests that our estimators possess desirable finite sample properties.

[7] Both BFGS and NR optimization methods are available as options in our `R` and Stata procedures. Using multistarts proved to work well both in our simulations and empirical examples.

3.3 Stata and R packages

All the analysis above can be performed using packages we have created in `R` (the `snreg` `R` package) and Stata statistical softwares. The `R` package and the Stata command can be obtained from the authors' websites. Both softwares are accompanied by help and example files. In both softwares the names of the commands are `snreg` and `snsf`. Different from the `selm` command from the `R` package `sn`, the `snreg` command allows for determinants of heteroskedasticity as in (14) and skewness as in (16). Appendix D presents `R` code to help replicate our empirical results, which we discuss next.

4 Empirical Illustration

In this section, we demonstrate the usefulness of our proposed methodology in three separate applications. We will look at both cost and production functions with symmetric and asymmetric noise. We will further introduce inefficiency of production units into our models. Finally, we will highlight our most flexible model that allows asymmetric noise and inefficiency, as well as determinants of (i) heteroskedasticity, (ii) inefficiency, and (iii) skewness.

We will showcase such comparisons by modeling risk in the U.S. banking industry, the effect of extreme adverse events on educational outcomes, and finally, annual data from the U.S. textile sector.

4.1 U.S. Banks

For our first application, we use a random subset of the firms employed in Restrepo-Tobón and Kumbhakar (2014). We chose a random sample of 500 banks observed in 2007 and whose total assets were between the 10th and 90th percentiles of the total assets distribution, and whose total costs were between the 10th and 90th percentiles of the total costs distribution.

The code to obtain our random sample is shown in Appendix **??**.[8]

Our goal is to estimate and compare the following models: (N0) symmetric noise with no inefficiency, (SN0) asymmetric noise with no inefficiency, (SF0) symmetric noise with inefficiency, (SF1) asymmetric noise with inefficiency and (SF2) asymmetric noise with inefficiency and determinants. These models go from the most restrictive to the most general. If the noise is asymmetric, inefficiency exists and our determinants are significant, we expect SF2 to perform best. However, if none of those events are true, N0 represents the most efficient model.

4.1.1 Translog Cost Function

We assume a full translog specification of the technology where 2 outputs are produced by 3 inputs. To ensure the necessary condition that the cost function is homogeneous of degree 1, we divide the total costs and prices of the first two inputs by the price of the third input.[9] More formally, our translog cost function is given as

$$
\begin{aligned}
\ln(TC/W_3) &= \beta_0 + \beta_1 \ln(Y_1) + \beta_2 \ln(Y_2) + \beta_3 \ln(W_1/W_3) + \beta_4 \ln(W_2/W_3) \\
&+ 0.5\beta_5 \ln(Y_1)^2 + 0.5\beta_6 \ln(Y_2)^2 + 0.5\beta_7 \ln(W_1/W_3)^2 \\
&+ 0.5\beta_8 \ln(W_2/W_3)^2 + \beta_9 \ln(Y_1)\ln(Y_2) + \beta_{10} \ln(Y_1)\ln(W_1/W_3) \\
&+ \beta_{11} \ln(Y_1)\ln(W_2/W_3) + \beta_{12} \ln(Y_2)\ln(W_1/W_3) \\
&+ \beta_{13} \ln(Y_2)\ln(W_2/W_3) + \beta_{14} \ln(W_1/W_3)\ln(W_2/W_3) + \epsilon
\end{aligned}
$$

where TC represents total costs of the bank, Y_1 and Y_2 are their outputs (total securities of the bank and total loans, respectively) and W_1, W_2 and W_3 are their inputs (cost of fixed assets, cost of labor and cost of borrowed funds, respectively).[10] Each of the β represent parameters

8 We repeated this experiment several times to ensure that the general conclusions were not dependent upon this particular sample of banks.

9 The choice of which input price is a numeraire does not affect the estimation.

10 For a more detailed description of the data, see Koetter et al. (2012).

to be estimated and the form of ϵ will depend upon the model chosen.

Table 1 presents the results of our translog cost function for each of the above specifications. Recall that Model N0 is the traditional cost function where the noise is homoskedastic and symmetric, i.e., $v_i \sim N(0, \sigma_v)$ in (3).[11] Model SN0 allows the noise to be SN, where the skewness parameter α is the same for all observations. Model SF0 is the standard SF model where noise is normally distributed with a constant variance and inefficiency is exponentially distributed. Model SF1 extends model SF0 by allowing noise to be SN.

Following the above discussion, we suggest a model where risk influences total costs of production through the noise. More specifically, for each bank, the shape/skewness of the distribution of the noise[12] depends upon the risk level of that bank. Thus, risk affects the total costs of a bank, not directly, but rather through the expected shock that a bank experiences due to being risky. Therefore, model SF2 allows the skewness parameter to be bank-specific as in (16). Here we employ a commonly used risk measure in the banking literature (Koetter et al., 2012), standard deviation of return on assets (*sdroa*).[13] It can be viewed as the variability in returns.

Model SF2 also adds explanatory variables for heteroskedasticity and inefficiency (Equations 14 and 15, respectively). For the variance, we look at the total assets (TA) of the bank and for inefficiency, we use a *scope* variable, which is the Hirschman-Herfindahl index across five loan categories (i.e., how focused a bank is in terms of loans).[14]

4.1.2 Results

Our most basic comparison is between the first two models: N0 and SN0 (symmetric and asymmetric noise without inefficiency, respectively). The estimated skewness coefficient is

---

[11] Model N0 is essentially OLS, however it is fit by the ML estimator under the assumption that the noise is normally distributed.

[12] The variance of noise is sometimes thought of as production risk (Just and Pope, 1978; Chavas et al., 2010).

[13] We also tried Z-score of a bank. These results are similar and are available upon request.

[14] The five categories of loans are listed as agricultural, commercial and industrial, individual, real estate, and other.

Table 1: Dependent variable $\ln(TC/W_3)$. z-values in parentheses.

Variable	N0	SN0	SF0	SF1	SF2
Intercept	-1.5833	-1.6923	-2.3203	-3.2175	-3.7655
	(-0.73)	(-0.80)	(-1.11)	(-1.54)	(-1.83)
ln(Y1)	0.3729	0.3753	0.3613	0.3271	0.2914
	(2.65)	(2.60)	(2.45)	(2.25)	(1.99)
ln(Y2)	0.4935	0.5738	0.7050	0.8497	0.9413
	(1.60)	(1.90)	(2.36)	(2.89)	(3.21)
ln(W1/W3)	0.3415	0.3292	0.3291	0.3133	0.3175
	(1.36)	(1.32)	(1.35)	(1.33)	(1.36)
ln(W2/W3)	-1.1573	-1.4004	-1.5193	-1.3617	-1.2350
	(-1.68)	(-2.10)	(-2.33)	(-2.07)	(-1.88)
0.5*ln(Y1)^2	0.0631	0.0586	0.0574	0.0598	0.0621
	(9.10)	(8.66)	(8.31)	(8.16)	(7.85)
0.5*ln(Y2)^2	0.0997	0.0890	0.0799	0.0730	0.0689
	(3.91)	(3.52)	(3.14)	(2.91)	(2.79)
0.5*ln(W1/W3)^2	-0.0577	-0.0561	-0.0561	-0.0567	-0.0507
	(-1.90)	(-1.87)	(-1.92)	(-1.94)	(-1.83)
0.5*ln(W2/W3)^2	0.8407	0.8471	0.8572	0.8530	0.8802
	(3.94)	(4.13)	(4.28)	(4.17)	(4.33)
ln(Y1)∗ln(Y2)	-0.0720	-0.0704	-0.0713	-0.0724	-0.0719
	(-5.58)	(-5.41)	(-5.37)	(-5.58)	(-5.48)
ln(Y1)∗ln(W1/W3)	-0.0239	-0.0265	-0.0277	-0.0290	-0.0290
	(-1.72)	(-1.88)	(-1.94)	(-2.13)	(-2.14)
ln(Y1)∗ln(W2/W3)	0.0183	0.0289	0.0435	0.0526	0.0551
	(0.52)	(0.81)	(1.24)	(1.58)	(1.61)
ln(Y2)∗ln(W1/W3)	0.0105	0.0129	0.0133	0.0142	0.0157
	(0.60)	(0.75)	(0.77)	(0.85)	(0.94)
ln(Y2)∗ln(W2/W3)	-0.0723	-0.0634	-0.0692	-0.0900	-0.1073
	(-1.54)	(-1.36)	(-1.50)	(-1.97)	(-2.36)
ln(W1/W3)∗ln(W2/W3)	-0.0376	-0.0361	-0.0332	-0.0257	-0.0369
	(-0.63)	(-0.61)	(-0.58)	(-0.46)	(-0.67)
$\ln\sigma_v^2$					
Intercept	-3.3997	-2.8624	-3.8485	-3.2090	0.1264
	(-53.75)	(-22.01)	(-33.09)	(-22.36)	(0.08)
ln(TA)					-0.2802
					(-2.14)
$\alpha(z)$					
Intercept		1.3822		-2.6274	-3.7828
		(4.90)		(-2.40)	(-2.59)
sdroa					3.9765
					(2.25)
$\ln\sigma_u^2$					
Intercept			-4.4483	-4.1864	-3.5314
			(-17.11)	(-19.34)	(-6.71)
scope					-1.6177
					(-1.94)
Log-likelihood	140.447	144.692	149.188	152.723	162.212

1.38, which is significant at conventional levels. The noise distribution is close to the pink density shown in Figure 1, mirrored around 0. The N0 model is rejected by the LR test in favor of the SN0 model (p-value of the LR test is 0.0036). Although OLS is unbiased, it is no longer efficient in the presence of asymmetric noise.

We now move to introducing inefficiency into our cost function.[15] Table 1 shows that the symmetric SF model SF0 exhibits better fit than SN0 with the same number of parameters. We should be careful here however as SF0 and SN0 are non-nested and hence the LR-test is not necessarily informative. When we allow both skewed noise and inefficiency (model SF1), the LR test clearly rejects SN0 in favor of SF1 (p-value of the LR test is 6.13e-05)[16] and also for SF1 in favor of SF0 (p-value of the LR test is 0.0078). However, note that SF1 restricts the shapes of the noise and inefficiency distributions to be the same for all banks. The most flexible model, SF2, best fits the data among all those considered in Table 1. The LR test gives preference to SF2 over SF1 (the p-value of the LR test is 7.57e-05).[17]

Figure 3 shows the kernel estimated density of the predicted skewness ($\hat{\alpha}(z)$) for our preferred model, SF2. The probability mass of a negatively skewed distribution with a zero

[15] Recall that with a symmetric noise such as in SF0, ϵ in (5) is negatively skewed for a production function and positively skewed for a cost function.

[16] The careful reader will have noticed that the signs of the skewness parameters in SN0 and SF1 are flipped. Note that they are not expected to have the same sign, as the noise in SF1 is only a part of the compound error term. The cumulant of noise in an SN model is

$$K_{v_{SN0}}(t) = \ln 2 - \sigma_{v_{SN0}} \sqrt{\frac{2}{\pi}} \frac{\alpha_{SN0}}{\sqrt{1+\alpha_{SN0}^2}} t + \frac{\sigma_{v_{SN0}}^2 t^2}{2} + \ln\left[\Phi\left(\sigma_{v_{SN0}} \frac{\alpha_{SN0}}{\sqrt{1+\alpha_{SN0}^2}} t\right)\right],$$

while the cumulant of noise in an SN-Exp model is

$$\begin{aligned} K_{\epsilon_{SF1}}(t) = \ln 2 - \sigma_{v_{SF1}} \sqrt{\frac{2}{\pi}} \frac{\alpha_{SF1}}{\sqrt{1+\alpha_{SF1}^2}} t + \frac{\sigma_{v_{SF1}}^2 t^2}{2} \\ + \ln\left[\Phi\left(\sigma_{v_{SF1}} \frac{\alpha_{SF1}}{\sqrt{1+\alpha_{SF1}^2}} t\right)\right] + \mathsf{p} \ln(1-\sigma_u t), \forall \sigma_u < 1/t. \end{aligned} \tag{17}$$

Even if the noise in both $SN0$ and $SF1$ have (roughly) the same third moment ($t = 3$), the α parameters are likely to be different due to presence of σ_u.

[17] In all LR tests, we consider a standard χ^2 distribution for the LR statistic as the tested parameters are not bounded.

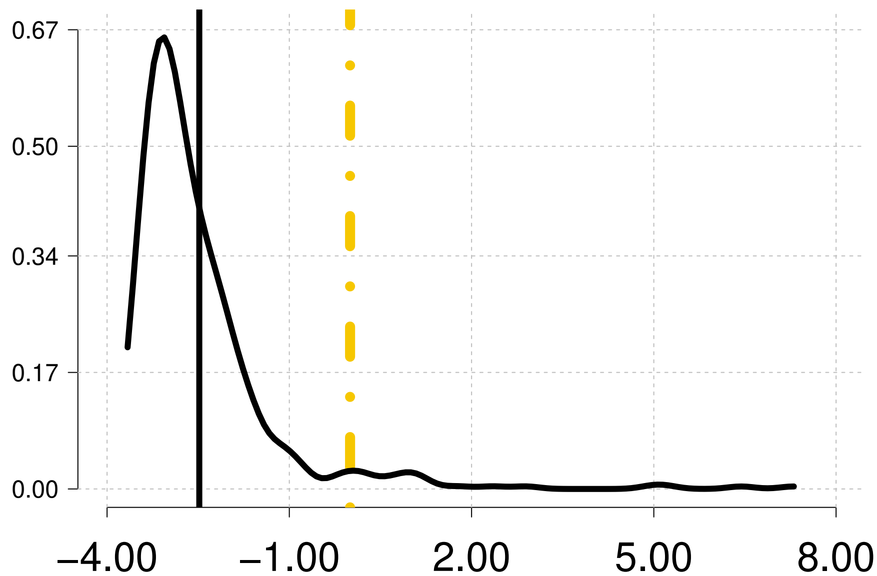


Figure 3: Kernel estimated densities of skewness (Model SF2). The vertical dash-dotted line is 0. The solid vertical line is the mean.

mean implies that the majority of banks are expected to have a slight negative shock to their operations. Another property of this distribution is that the left tail is thicker than the right. In other words, large positive shocks are more frequent than large negative shocks.

Figure 4 plots the predicted skewness against a skewness determinant. There are only a few very risky banks (i.e., *sdroa* is very large). At low risk levels, the skewness is quite low (approximately -4 for *sdroa*). A shock of a low risk bank comes from a very skewed distribution and therefore such a bank is likely to be hit by a negative shock that has a detrimental effect on total costs. At the mean level of *sdroa* (0.32), the estimated skewness is -2.5, the value of the estimated skewness in SF1 (where skewness is assumed constant). The skewness remains negative until *sdroa* reaches 0.95, which is the 96th percentile of the *sdroa* distribution. For the 4 percent (of the most) risky banks in our sample, the skewness is positive, implying a thicker right tail of the noise distribution.

It is worth noting that in SF2, the inefficiency determinant *scope*, is statistically significant. The negative coefficient means that as *scope* increases, bank inefficiency is decreasing. Further, the determinant of heteroskedasticity (total assets) is also statistically significant. The negative coefficient here suggests that the variance decreases with the size of total assets.

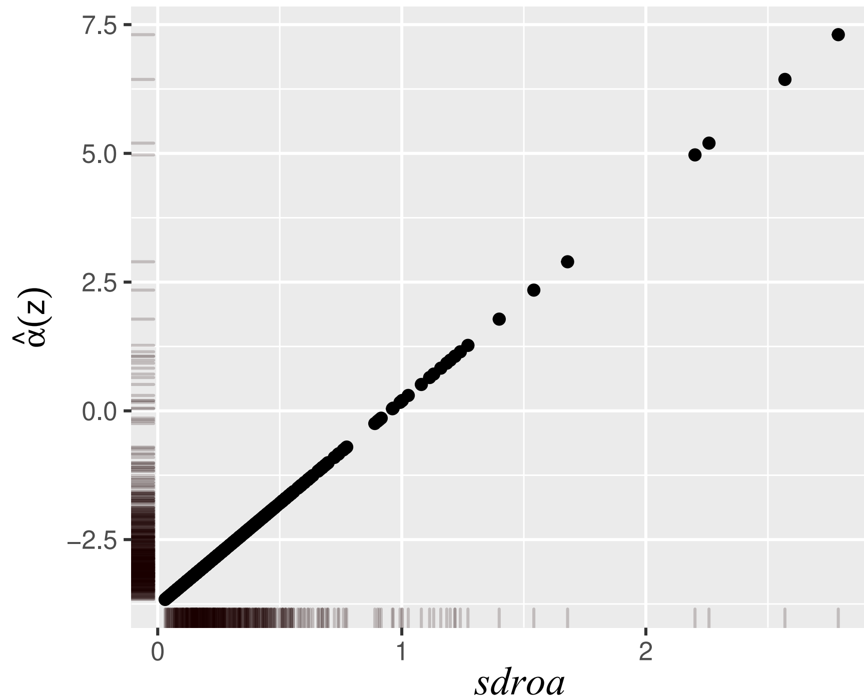


Figure 4: The estimate of skewness (fitted values) plotted against the determinant (Model SF2). The rug plot on each axis essentially shows a one-dimensional heatmap.

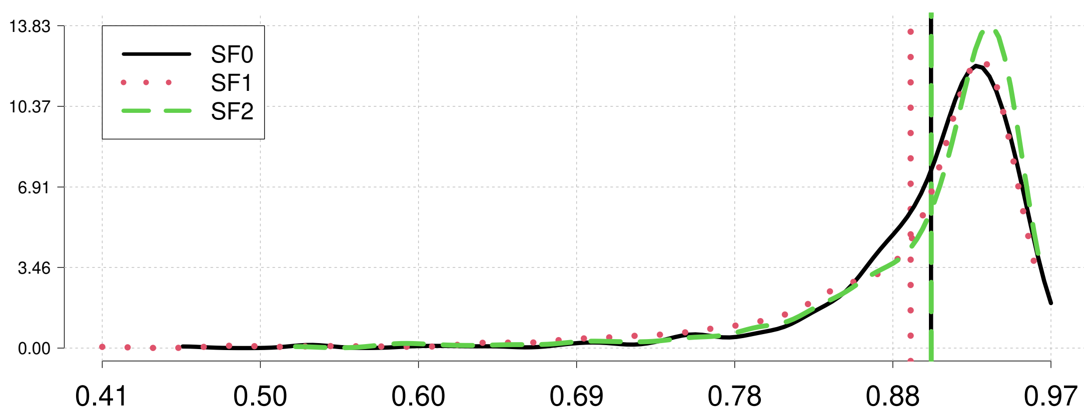


Figure 5: Kernel estimated densities of efficiencies. Vertical lines are respective means.

Finally, Figure 5 shows estimated densities of efficiency scores from all of our SF models. There are no marked differences in the distributions.

4.2 Beltway Sniper

Here we investigate the effects of the 2002 "Beltway Sniper" mass shootings on student achievement in Virginia's public elementary schools (Gershenson and Tekin, 2018). Traumatic events, especially those which are 'close to home', can have serious impacts on student outcomes. However, different from past research (Ponzo, 2013), we attempt to model these low probability events in the noise distribution.[18]

We follow Levin (1974) and Hanushek (1979) and consider an educational production function as a process of converting inputs (i.e., school resources) into outputs (i.e., student achievement). We go a step further and account for inefficiency in educational production as it has been argued that estimating educational production functions accounting for inefficiency is a proper approach for examining educational outcomes (Thanassoulis et al., 2016, 2018; Ruggiero, 2006, 2019).

4.2.1 Educational Production Function

Our (school level) educational production function is given as

$$\begin{aligned}\ln(math) &= \beta_0 + \beta_1 \ln(ratio) + \beta_2 \ln(fte) + 0.5\beta_3 \ln(ratio)^2 \\ &+ 0.5\beta_4 \ln(fte)^2 + \beta_5 \ln(ratio)\ln(fte) \\ &+ \beta_6 \ln(black) + \beta_7 \ln(hispanic) + \epsilon,\end{aligned}$$

where the βs represent parameters to be estimated and the composition of ϵ follows the same models in the previous sub-section. *math* is our output variable measured in logs (school-level proficiency in the Standards of Learning standardized test given each spring in Virginia public schools). Our input variables are student-teacher ratios (*ratio*), full-time equivalent teachers (*fte*), percent black (*black*), and percent Hispanic (*hispanic*).

[18] It would be interesting to see our estimator applied to studies about the effect of bullying on educational outcomes (e.g., Lacey and Cornell, 2013).

Similar to before, we estimate five different models (N0, SN0, SF0, SF1 and SF2). In this context, “production inefficiency” represents student underachievement. We will use total enrollment (*enroll*) as a measure of size, percent free lunch (frp), and closeness (*closeness*) to a sniper attack as determinants of our noise components.[19] *closeness* is the primary determinant of interest and measures the distance (in miles) the school is from the closest sniper attack.

4.2.2 Results

Table 2 provides the regression results for our familiar set of models. Note that in the previous sub-section we analyzed a cost function (i.e., $\mathsf{p} = -1$) and the smaller outcome variable was preferable. Here, we analyzing a production function (i.e., $\mathsf{p} = 1$) and larger outcome values are preferable (i.e., higher levels of proficiency). The results here represent 5th grade students in the year 2003 (same academic year as the attacks).

It is clear that the SN0 model fits the data far better than N0 (the LR statistic is 136.53 while the critical value of the χ^2_1 at the 1% level of significance is 6.63), and thus there is a good reason to believe the skewness of the noise term is not 0. Based on the LR test, including inefficiency (SF0) provides a better fit than simply allowing for asymmetric noise (the LR statistic is 40.97). When we consider the model that contains both inefficiency and skewness, restricting the shape of the noise distribution for all schools to be the same (model SF1), the constant skewness parameter is not statistically insignificant. The likelihood increased by only 0.4, which is not enough to conclude that SF1 is preferred to SF0. Note that when we do not account for possible skewness in the noise, we overestimate the effect of the proportion of black or Hispanic students on educational outcome.

The most flexible model (SF2) allows the skewness parameter to vary depending on how close the school is from a shooting scene. We find a significant increase in the log-likelihood (the LR statistic of the LR test between SF2 and SF1 is 118.4 whereas the critical value

[19] Estimating different specifications of an educational production function and noise components led to the same conclusions.

Table 2: Dependent variable is log of percent proficient of the Math test. z-values in parentheses.

	Dep. var is log ($math$)				
Variable	N0	SN0	SF0	SF1	SF2
Intercept	4.0145	4.0140	4.3207	4.3029	3.9931
	(5.17)	(4.16)	(10.62)	(11.69)	(6.90)
ln(fte)	0.1850	0.1516	0.1204	0.1219	0.2576
	(0.74)	(0.52)	(0.88)	(0.96)	(1.47)
ln($ratio$)	0.2056	0.1579	0.1344	0.1453	0.2415
	(0.69)	(0.42)	(0.85)	(1.01)	(1.08)
ln(fte)^2	-0.0184	-0.0110	-0.0124	-0.0130	-0.0274
	(-0.90)	(-0.53)	(-1.10)	(-1.20)	(-2.07)
ln($ratio$)^2	-0.0209	-0.0135	-0.0144	-0.0162	-0.0253
	(-0.65)	(-0.35)	(-0.84)	(-1.00)	(-1.12)
ln(fte)*ln($ratio$)	-0.0389	-0.0368	-0.0264	-0.0261	-0.0544
	(-0.86)	(-0.65)	(-1.05)	(-1.12)	(-1.63)
$black$	-0.4519	-0.2373	-0.2291	-0.2299	-0.1330
	(-13.21)	(-9.08)	(-9.63)	(-9.66)	(-5.01)
$hispanic$	-0.4240	-0.3376	-0.3569	-0.3627	-0.1872
	(-6.05)	(-5.68)	(-7.53)	(-7.63)	(-3.33)
$\ln \sigma_v^2$					
Intercept	-3.3138	-2.4067	-5.5637	-5.1480	-4.7220
	(-53.18)	(-44.16)	(-24.87)	(-16.86)	(-12.59)
$enroll$					0.0016
					(3.12)
$\alpha(z)$					
Intercept		-6.4106		1.4828	-11.3756
		(-5.67)		(1.24)	(-2.38)
ln($closeness$)					2.7476
					(2.07)
$\ln \sigma_u^2$					
Intercept			-3.3859	-3.3468	-6.1308
			(-27.37)	(-25.72)	(-14.70)
frp					5.7394
					(8.71)
N	515	515	515	515	515
Log-likelihood	122.544	190.810	211.297	211.679	270.878

of the χ_3^2 at the 1% level of significance is 11.34). As for the determinants of the error components, we find that as the proportion of pupils who are eligible for free or reduced-price

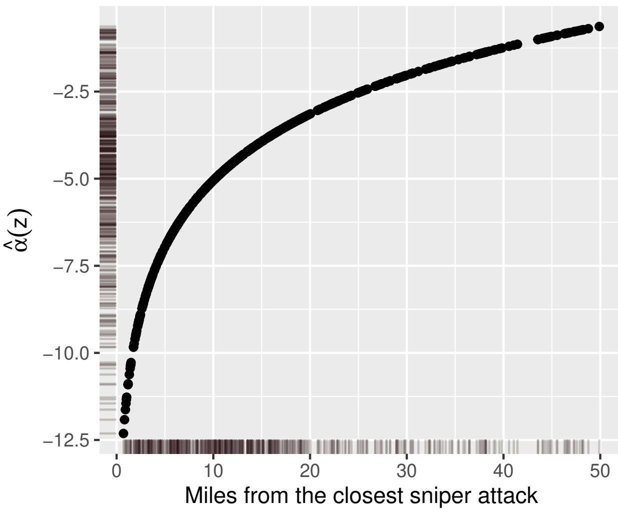


Figure 6: The estimate of skewness (fitted values) plotted against the respective determinant. The rug plot on each axis essentially shows a one-dimensional heatmap.

lunch is increasing, underachievement is increasing. The skewness of the noise distribution is increasing as a school is further away from a sniper attack (as shown in the Figure 6).[20] The noise for those schools that are close to at least one sniper attack scene, have large negative skewness, implying that the left tail is much thicker than the right tail. In other words, as risk is increasing, schools have a larger probability to exhibit poor, rather than good test results. Figure 7 shows that negative skewness is a feature of the noise distribution for all public schools in our sample.

4.3 NBER Data: Textile Industries

Our final application uses data from the well studied (Bonanno et al., 2017) NBER-CES Manufacturing Industry Database (Bartelsman and Gray, 1996). For each available year (1958 – 2011), we focus on the textile industry (SIC 4-digit industry: 2200 – 2399) because these particular samples are known to exhibit the "wrong skewness" of OLS residuals (Hafner et al., 2018).

[20] As in the previous application, we find a positive relationship between the determinant and skewness parameter. However, in this application, a larger determinant implies a lower risk.

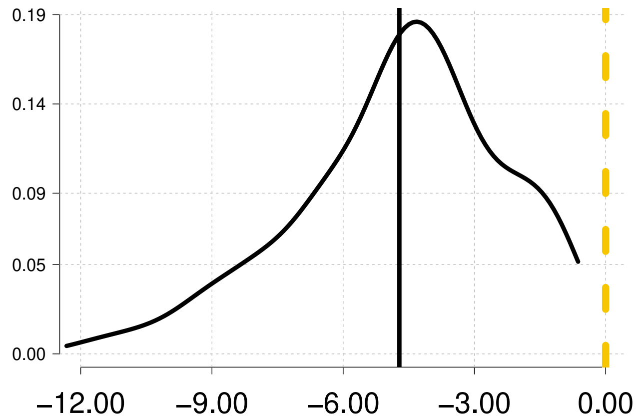


Figure 7: Kernel estimated density of skewness. The vertical dashed-dotted line is 0. The solid vertical line is the mean.

We estimate SF models where noise is either assumed to normal or SN and the distribution of the inefficiency term is assumed to be exponentially distributed. In the case of a SN distribution, we omit determinants and therefore have a constant skewness parameter for each year. In each setting, we use a translog production function where the output (total value added) is produced by capital (total real capital stock), labor (total employment) and materials (total cost of materials).

Figure 8 plots the estimated skewness parameter for each year. The blue circles represent coefficients that are statistically insignificant, while the red triangles represent statistically significant estimates of α. We do not observe uniformity of coefficient magnitudes; they range from roughly -2.5 to approximately 6. Although we see both signs for skewness, most estimates are close to 0. With regards to the magnitude, there is no clustering, trend, or situation where the estimates appear to be persistent over time. The skewness coefficient can be negative in one year and positive the year after. Finally, there appears to be no clustering or trend with respect to significance of the estimated coefficients.

Figure 9 dissects Figure 8 to differentiate between years where the skewness of the OLS residuals are negative (‘correct’ skewness) or positive (‘wrong’ skewness). In years where the SN-Exp model results in a large positive significant skewness parameter, the skewness of OLS residuals is ‘wrong’. Where the skewness of OLS residuals is ‘correct’, the skewness parameter

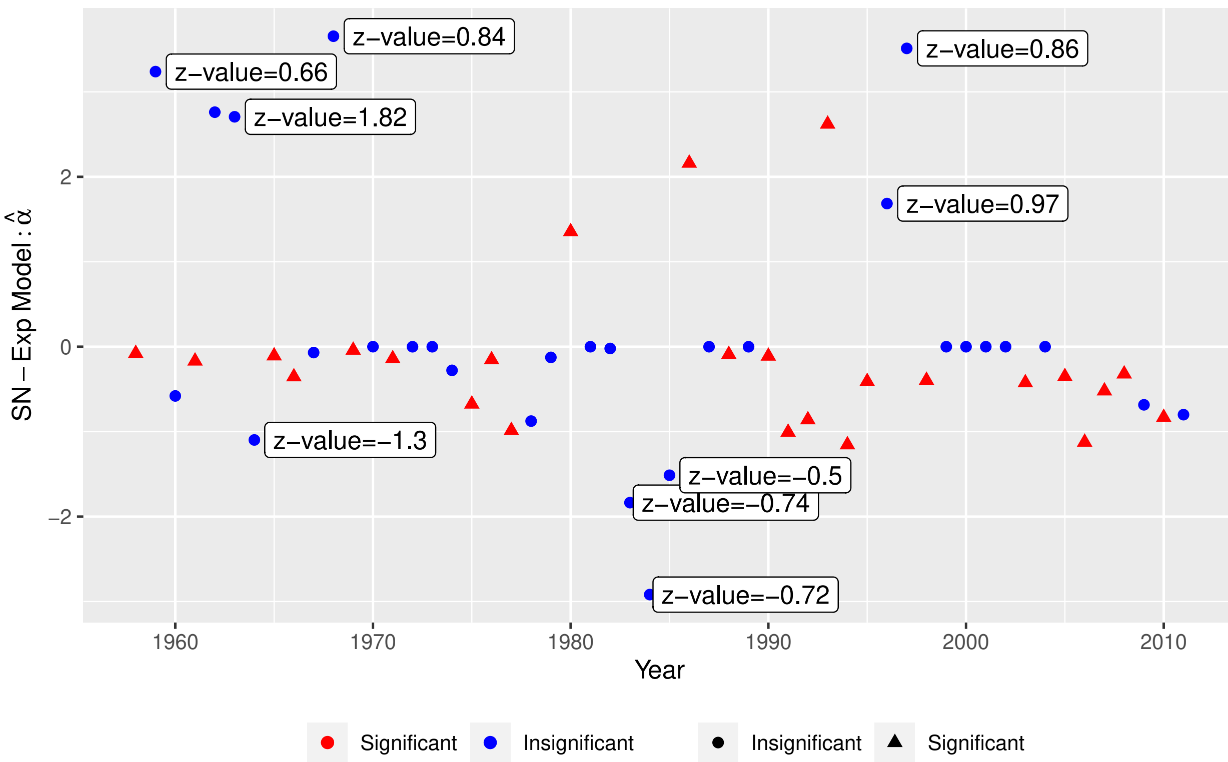


Figure 8: The estimated skewness coefficient by year.

is only rarely significant.

Finally, Figure 10 shows average efficiency scores by year for both the asymmetric and symmetric noise models. In years where the OLS residuals are of the 'wrong skewness', the conventional SF model predicts no inefficiency (i.e., the average efficiency score is 1). We observe this in about half of the cases. In each of those years, our model estimates inefficiency (the average efficiency score is around 0.975). In about 10% of the years, the average efficiencies from both models are the same. This happens in years when the skewness of the OLS residuals is 'correct' and the estimated skewness coefficient in our model is indistinguishable from 0 and is statistically insignificant (red circles in Figure 9). Considering both what we have seen here and our simulations, there is evidence that our approach can identify inefficiency in each year of our sample.

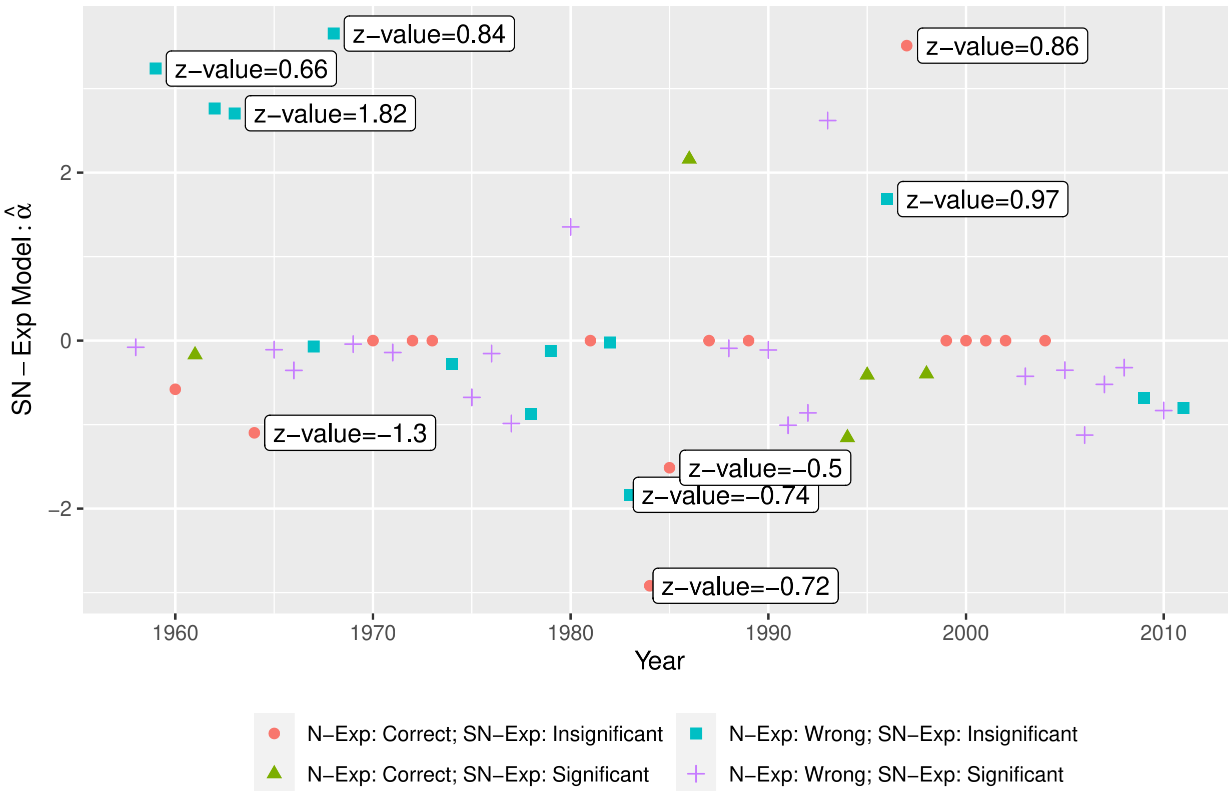


Figure 9: The estimated skewness coefficient by year.

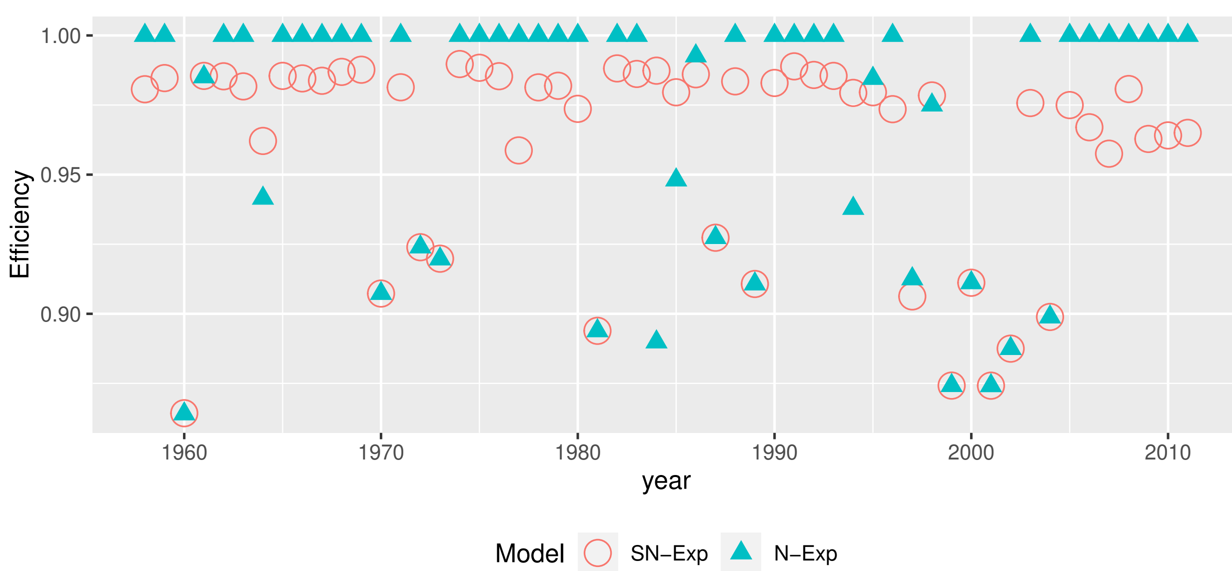


Figure 10: The average of the efficiency score estimated by N-Exp and SN-Exp models by year.

5 Conclusions

In this paper, we propose to model asymmetric noise in production analysis. We discussed how to estimate a production or cost function with asymmetric noise and extended this model for a skew-normal noise distribution for stochastic frontier analysis. Our methods result in closed form solutions for the log-likelihood function and inefficiency. We are able to incorporate determinants of these components (heteroskedasticity, inefficiency and skewness) in an estimation procedure that jointly estimates all parameters of interest.

We showcased these methods in simulations as well as in three separate empirical applications, including one that showed that our approach is able to estimate efficiency scores when OLS residuals are of the "wrong skewness". Given that we have produced user-friendly `R` and Stata packages, we believe that these techniques can easily be applied across a wide range of fields within production analysis.

References

Adcock, C.J., 2005. Exploiting skewness to build an optimal hedge fund with a currency overlay. The European Journal of Finance 11, 445–462.

Adcock, C.J., 2007. Extensions of Stein's Lemma for the Skew-Normal Distribution. Communications in Statistics - Theory and Methods 36, 1661–1671.

Adcock, C.J., 2010. Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. Annals of Operations Research 176, 221–234.

Afriat, S.N., 1972. Efficiency Estimation of Production Functions. International Economic Review 13, 5–23.

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

Amsler, C., Prokhorov, A., Schmidt, P., 2016. Endogeneity in stochastic frontier models. Journal of Econometrics 190, 280–288.

Aubuchon, C.P., Wheelock, D.C., 2010. The Geographic Distribution and Characteristics of

U.S. Bank Failures, 2007-2010: Do Bank Failures Still Reflect Local Economic Conditions? Federal Bank of St. Louis Review 92, 395–415.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61, 579–602.

Azzalini, A., Capitanio, A., 2013. The Skew-Normal and Related Families. Institute of Mathematical Statistics Monographs, Cambridge University Press.

Badunenko, O., Henderson, D.J., Kumbhakar, S.C., 2012. When, where and how to perform efficiency estimation. Journal of the Royal Statistical Society: Series A (Statistics in Society) 175, 863–892.

Badunenko, O., Kumbhakar, S.C., 2016. When, where and how to estimate persistent and transient efficiency in stochastic frontier panel data models. European Journal of Operational Research 255, 272–287.

Ball, L., Mankiw, N.G., 1995. Relative-Price Changes as Aggregate Supply Shocks. The Quarterly Journal of Economics 110, 161–193.

Bartelsman, E.J., Gray, W., 1996. The NBER Manufacturing Productivity Database. Working Paper 205. National Bureau of Economic Research.

Bonanno, G., De Giovanni, D., Domma, F., 2017. The 'wrong skewness' problem: A respecification of stochastic frontiers. Journal of Productivity Analysis 47, 49–64.

Caudill, S.B., Ford, J.M., Gropper, D.M., 1995. Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity. Journal of Business & Economic Statistics 13, 105–11.

Chavas, J.P., Chambers, R.G., Pope, R.D., 2010. Production Economics and Farm Management: a Century of Contributions. American Journal of Agricultural Economics 92, 356–375.

Cho, C.K., Schmidt, P., 2020. The wrong skew problem in stochastic frontier models when inefficiency depends on environmental variables. Empirical Economics 58, 2031–2047.

Day, R.H., 1965. Probability Distributions of Field Crop Yields. Journal of Farm Economics 47, 713–741.

DiCiccio, T.J., Monti, A.C., 2004. Inferential Aspects of the Skew Exponential Power Distribution. Journal of the American Statistical Association 99, 439–450.

Domínguez-Molina, J.A., González-Farías, G., Ramos-Quiroga, R., 2004. Coastal Flooding and the Multivariate Skew-t Distribution, in: Genton, M.G. (Ed.), Skew-Elliptical Distributions and Their Applications. 1st edition ed.. Chapman and Hall/CRC, New York. chapter 14, pp. 243–258.

Florens, J.P., Simar, L., Van Keilegom, I., 2020. Estimation of the Boundary of a Variable Observed With Symmetric Error. Journal of the American Statistical Association 115, 425–441.

Genton, M.G. (Ed.), 2004. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. 1st editio ed., Chapman and Hall/CRC, New York.

Gershenson, S., Tekin, E., 2018. The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks. Education Finance and Policy 13, 513–544.

Goh, J.W., Lim, K.G., Sim, M., Zhang, W., 2012. Portfolio value-at-risk optimization for asymmetrically distributed asset returns. European Journal of Operational Research 221, 397–406.

Greene, W.H., 2008. The Econometric Approach to Efficiency Analysis, in: The Measurement of Productive Efficiency and Productivity Change. Oxford University Press, New York.

Hafner, C.M., Manner, H., Simar, L., 2018. The "wrong skewness" problem in stochastic frontier models: A new approach. Econometric Reviews 37, 380–400.

Hanushek, E.A., 1979. Conceptual and empirical issues in the estimation of educational production functions. Journal of Human Resources 14, 351–388.

Horrace, W.C., 2005a. On ranking and selection from independent truncated normal distributions. Journal of Econometrics 126, 335–354. Current developments in productivity and efficiency measurement.

Horrace, W.C., 2005b. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis 94, 209–221.

Horrace, W.C., Parmeter, C.F., 2018. A laplace stochastic frontier model. Econometric Reviews 37, 260–280.

Johnes, J., Portela, M., Thanassoulis, E., 2017. Efficiency in education. Journal of the Operational Research Society 68, 331–338.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics 19, 233–238.

Jradi, S., Parmeter, C.F., Ruggiero, J., 2021. Quantile estimation of stochastic frontiers with the normal-exponential specification. European Journal of Operational Research .

Just, R.E., Pope, R.D., 1978. Stochastic Specification of Production Functions and Economic Implications. Journal of Econometrics 7, 67–86.

Koetter, M., Kolari, J.W., Spierdijk, L., 2012. Enjoying the Quiet Life Under Deregulation? Evidence from Adjusted Lerner Indicies for U.S. Banks. Review of Economics and Statistics 94, 462–480.

Koopmans, T.C., 1951. An Analysis of Production as an Efficient Combination of Activities, in: Koopmans, T.C. (Ed.), Activity Anlaysis of Production and Allocation. Cowles Commission for Research in Economics, New York: Wiley. chapter 13.

Kumbhakar, S.C., Ghosh, S., McGuckin, J.T., 1991. A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. Journal of Business & Economic Statistics 9, 279–286.

Kumbhakar, S.C., Tsionas, E.G., 2005. The Joint Measurement of Technical and Allocative Inefficiencies. Journal of the American Statistical Association 100, 736–747.

Lacey, A., Cornell, D., 2013. The Impact of Teasing and Bullying on Schoolwide Academic Performance. Journal of Applied School Psychology 29, 262–283.

Lai, H.p., Kumbhakar, S.C., 2018. Endogeneity in panel data stochastic frontier model with determinants of persistent and transient inefficiency. Economics Letters 162, 5–9.

Levin, H.M., 1974. Measuring efficiency in educational production. Public Finance Quarterly 2, 3–24.

Li, Q., 1996. Estimating a stochastic production frontier when the adjusted error is symmetric. Economics Letters 52, 221–228.

Lien, G., Kumbhakar, S.C., Alem, H., 2018. Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms. International Journal of Production Economics 201, 53–61.

Louis, H., Blenman, L.P., Thatcher, J.S., 1999. Interest Rate Parity and the Behavior of the Bid-Ask Spread. Journal of Financial Research 22, 189–206.

Mao, Z., Liu, W., Feng, B., 2019. Opaque distribution channels for service providers with asymmetric capacities: Posted-price mechanisms. International Journal of Production Economics 215, 112–120.

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

Olson, J.A., Schmidt, P., Waldman, D.M., 1980. A monte carlo study of estimators of stochastic frontier production functions. Journal of Econometrics 13, 67–82.

Owen, D.B., 1956. Tables for Computing Bivariate Normal Probabilities. The Annals of Mathematical Statistics 27, 1075–1090.

Owen, D.B., 1980. A table of normal integrals. Communications in Statistics - Simulation and Computation 9, 389–419.

Papadopoulos, A., 2021. Stochastic frontier models using the Generalized Exponential distribution. Journal of Productivity Analysis 55, 15–29.

Ponzo, M., 2013. Does bullying reduce educational achievement? an evaluation using matching estimators. Journal of Policy Modeling 35, 1057–1078.

Qi, L., Bravo-Ureta, B.E., Cabrera, V.E., 2015. From cold to hot: Climatic effects and productivity in Wisconsin dairy farms. Journal of Dairy Science 98, 8664–8677.

Restrepo-Tobón, D., Kumbhakar, S.C., 2014. Enjoying the Quiet Life Under Deregulation? Not Quite. Journal of Applied Econometrics 29, 333–343.

Ritter, C., Bates, D.M., 1996. Profile methods, in: Prat, A. (Ed.), COMPSTAT, Physica-Verlag HD, Heidelberg. pp. 123–134.

Ruggiero, J., 1996. Efficiency of educational production: An analysis of new york school districts. Review of Economics and Statistics 78, 499–509.

Ruggiero, J., 2006. Measurement error, education production and data envelopment analysis. Economics of Education Review 25, 327–333. Special Issue: In Honor of W. Pierce Liles.

Ruggiero, J., 2019. The Choice of Comparable DMUs and Environmental Variables. Springer International Publishing, Cham. pp. 123–144.

Schmidt, P., 2011. One-Step and Two-Step Estimation in SFA Models. Journal of Productivity Analysis 36, 201–203.

Shaik, S., 2013. Crop Insurance Adjusted Panel Data Envelopment Analysis Efficiency Measures. American Journal of Agricultural Economics 95, 1155–1177.

Simar, L., Wilson, P.W., 2009. Inferences from cross-sectional, stochastic frontier models. Econometric Reviews 29, 62–98.

Stead, A.D., Wheat, P., Greene, W.H., 2019. Distributional Forms in Stochastic Frontier Analysis. Springer International Publishing, Cham. pp. 225–274.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57–66.

Sun, K., Henderson, D.J., Kumbhakar, S.C., 2011. Biases in approximating log production. Journal of Applied Econometrics 26, 708–714.

Thanassoulis, E., De Witte, K., Johnes, J., Johnes, G., Karagiannis, G., Portela, C.S., 2016. Applications of Data Envelopment Analysis in Education. Springer US, Boston, MA. pp. 367–438.

Thanassoulis, E., Dey, P.K., Petridis, K., Goniadis, I., Georgiou, A.C., 2017. Evaluating higher education teaching performance using combined analytic hierarchy process and data envelopment analysis. Journal of the Operational Research Society 68, 431–445.

Thanassoulis, E., Sotiros, D., Koronakos, G., Despotis, D., 2018. Assessing the cost-effectiveness of university academic recruitment and promotion policies. European Journal of Operational Research 264, 742–755.

Thanssoulis, E., 1999. Setting achievement targets for school children. Education Economics 7, 101–119.

Wang, H.j., Schmidt, P., 2002. One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. Journal of Productivity Analysis 18, 129–144.

Wei, Z., Zhu, X., Wang, T., 2021. The extended skew-normal-based stochastic frontier model with a solution to 'wrong skewness' problem. Statistics , in press.

Wheat, P., Stead, A.D., Greene, W.H., 2019. Robust stochastic frontier analysis: a Student's t-half normal model with application to highway maintenance costs in England. Journal of Productivity Analysis 51, 21–38.

de Witte, K., López-Torres, L., 2017. Efficiency in education: a review of literature and a way forward. Journal of the Operational Research Society 68, 339–363.

de Witte, K., Thanassoulis, E., Simpson, G., Battisti, G., Charlesworth-May, A., 2010. Assessing pupil and school performance by non-parametric and parametric techniques. Journal of the Operational Research Society 61, 1224–1237.

Yao, F., Zhang, F., Kumbhakar, S.C., 2019. Semiparametric smooth coefficient stochastic frontier model with panel data. Journal of Business & Economic Statistics 37, 556–572.

Yao, W., Zhao, Z., 2013. Kernel density-based linear regression estimate. Communications in Statistics - Theory and Methods 42, 4499–4512.

Supplementary material

A Skew-Normal Exponential Stochastic Frontier Model

A.1 Integral

Before deriving the $f(\epsilon)$, consider the following integral due to Owen (1980),

$$\int \phi(\boldsymbol{x})\,\Phi(a+b\boldsymbol{x})\,d\boldsymbol{x} = T\left(\boldsymbol{x}, \frac{a}{\boldsymbol{x}\sqrt{1+b^2}}\right) + T\left(\frac{a}{\sqrt{1+b^2}}, \frac{\boldsymbol{x}\sqrt{1+b^2}}{a}\right) - T\left(\boldsymbol{x}, \frac{a+b\boldsymbol{x}}{\boldsymbol{x}}\right) - T\left(\frac{a}{\sqrt{1+b^2}}, \frac{ab+\boldsymbol{x}(1+b^2)}{a}\right) + \Phi(\boldsymbol{x})\,\Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \tag{A1}$$

Denote $a_2 = \dfrac{a}{\sqrt{1+b^2}}$, then (A1) becomes

$$\underbrace{T\left(\boldsymbol{x}, \frac{a_2}{\boldsymbol{x}}\right)}_{\text{term 1}} + \underbrace{T\left(a_2, \frac{\boldsymbol{x}}{a_2}\right)}_{\text{term 2}} - \underbrace{T\left(\boldsymbol{x}, b+\frac{a}{\boldsymbol{x}}\right)}_{\text{term 3}} - \underbrace{T\left(a_2, b+\frac{\boldsymbol{x}(1+b^2)}{a}\right)}_{\text{term 4}} + \underbrace{\Phi(\boldsymbol{x})\,\Phi(a_2)}_{\text{term 5}}. \tag{A2}$$

Integral (A2) is needed on the domain $[x_1, +\infty)$. We will use three properties of Owen's T function (see Owen, 1956) to simplify the integral in (A2). First,

$$T(H,A) + T\left(AH, \frac{1}{A}\right) = \begin{cases} \frac{1}{2}\left(\Phi(H)+\Phi(AH)\right) - \Phi(H)\Phi(AH) & \text{if} \quad A \geq 0 \\ \frac{1}{2}\left(\Phi(H)+\Phi(AH)\right) - \Phi(H)\Phi(AH) - \frac{1}{2} & \text{if} \quad A < 0 \end{cases}. \tag{A3}$$

Second,

$$T(H,0) = 0. \tag{A4}$$

Third,

$$T(H,+\infty) = \frac{1}{2}\Phi\left(-\left|H\right|\right). \tag{A5}$$

Integrate individual terms in (A2)

$$
\begin{aligned}
(\text{term 1})\Big|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} &= \underbrace{T\left(\boldsymbol{x}, \frac{a_2}{\boldsymbol{x}}\right)}_{\text{term 1}}\Bigg|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} \qquad (\text{A6})\\
&= T\left(+\infty, \frac{a_2}{+\infty}\right) - T\left(x_1, \frac{a_2}{x_1}\right) = -T\left(x_1, \frac{a_2}{x_1}\right),\\
(\text{term 2})\Big|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} &= \underbrace{T\left(a_2, \frac{\boldsymbol{x}}{a_2}\right)}_{\text{term 2}}\Bigg|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} \qquad (\text{A7})\\
&= T\left(a_2, +\infty\right) - T\left(a_2, \frac{x_1}{a_2}\right),\\
(\text{term 3})\Big|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} &= \underbrace{-T\left(\boldsymbol{x}, b + \frac{a}{\boldsymbol{x}}\right)}_{\text{term 3}}\Bigg|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} \qquad (\text{A8})\\
&= -T\left(+\infty, b + \frac{a}{+\infty}\right) + T\left(x_1, b + \frac{a}{x_1}\right) = T\left(x_1, b + \frac{a}{x_1}\right),\\
(\text{term 4})\Big|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} &= \underbrace{-T\left(a_2, b + \frac{\boldsymbol{x}(1+b^2)}{a}\right)}_{\text{term 4}}\Bigg|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} \qquad (\text{A9})\\
&= -T\left(a_2, +\infty\right) + T\left(a_2, b + \frac{x_1(1+b^2)}{a}\right),\\
(\text{term 5})\Big|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} &= \underbrace{\Phi\left(\boldsymbol{x}\right)\Phi\left(a_2\right)}_{\text{term 5}}\Bigg|_{\boldsymbol{x}=x_1}^{\boldsymbol{x}=+\infty} \qquad (\text{A10})\\
&= \Phi\left(+\infty\right)\Phi\left(a_2\right) - \Phi\left(x_1\right)\Phi\left(a_2\right) = \Phi\left(a_2\right) - \Phi\left(x_1\right)\Phi\left(a_2\right).
\end{aligned}
$$

Adding these 5 terms yields the closed form solution to the integral in (A2):

$$
\begin{aligned}
\int_{x_1}^{\infty} \phi\left(\boldsymbol{x}\right)\Phi\left(a + b\boldsymbol{x}\right) d\boldsymbol{x} = {} & \underbrace{-T\left(x_1, \frac{a_2}{x_1}\right)}_{\text{term 1}} \underbrace{-T\left(a_2, \frac{x_1}{a_2}\right)}_{\text{term 2}} + \underbrace{T\left(x_1, b + \frac{a}{x_1}\right)}_{\text{term 3}} + \\
& \underbrace{T\left(a_2, b + \frac{x_1(1+b^2)}{a}\right)}_{\text{term 4}} + \underbrace{\Phi\left(a_2\right)\Phi\left(-x_1\right)}_{\text{term 5}}. \qquad (\text{A11})
\end{aligned}
$$

We can leave it as is or we can simplify it further by using the first property (A3) of Owen's

T function. Denote $H = x_1$ and $A = \dfrac{a_2}{x_1}$ and consider the first two terms in (A11). Then

$$\begin{aligned}
&\underbrace{-T\left(x_1, \frac{a_2}{x_1}\right)}_{\text{term 1}} - \underbrace{T\left(a_2, \frac{x_1}{a_2}\right)}_{\text{term 2}} && \text{(A12)}\\
&= -\frac{1}{2}\left[\Phi(H) + \Phi(AH)\right] + \Phi(H)\Phi(AH) + \frac{1}{2} I\left(A < 0\right)\\
&= -\frac{1}{2}\left[\Phi(x_1) + \Phi\left(a_2\right)\right] + \Phi(x_1)\Phi\left(a_2\right) + \frac{1}{2} I\left(\frac{a_2}{x_1} < 0\right). && \text{(A13)}
\end{aligned}$$

Then the sum of term 1, term 2, and term 5 in (A11) is given by

$$\begin{aligned}
&\underbrace{-T\left(x_1, \frac{a_2}{x_1}\right)}_{\text{term 1}} - \underbrace{T\left(a_2, \frac{x_1}{a_2}\right)}_{\text{term 2}} + \underbrace{\Phi\left(a_2\right) - \Phi\left(x_1\right)\Phi\left(a_2\right)}_{\text{term 5}} && \text{(A14)}\\
&= -\frac{1}{2}\Phi(x_1) - \frac{1}{2}\Phi\left(a_2\right) + \Phi(x_1)\Phi\left(a_2\right) + \frac{1}{2} I\left(\frac{a_2}{x_1} < 0\right) + \Phi\left(a_2\right) - \Phi\left(x_1\right)\Phi\left(a_2\right)\\
&= -\frac{1}{2}\Phi(x_1) + \frac{1}{2}\Phi\left(a_2\right) + \frac{1}{2} I\left(\frac{a_2}{x_1} < 0\right). && \text{(A15)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\int_{x_1}^{\infty} \phi\left(\boldsymbol{x}\right)\Phi\left(a + b\boldsymbol{x}\right)d\boldsymbol{x} = {}& \underbrace{T\left(x_1, b + \frac{a}{x_1}\right)}_{\text{term 3}} + \underbrace{T\left(a_2, b + \frac{x_1(1+b^2)}{a}\right)}_{\text{term 4}}\\
&+ \underbrace{\frac{1}{2}\left[\Phi\left(a_2\right) - \Phi\left(x_1\right) + I\left(\frac{a_2}{x_1} < 0\right)\right]}_{\text{term 1 + term 2 + term 5}}. && \text{(A16)}
\end{aligned}$$

A.2 Derivation of $f(\epsilon)$ and the Log-Likelihood Function

The model is

$$y = f(\boldsymbol{x}; \boldsymbol{\beta}) + v - \mathsf{p}u = f(\boldsymbol{x}; \boldsymbol{\beta}) + \epsilon, \tag{A17}$$

in which the inefficiency term is exponentially distributed, so its density is given by

$$f_u(u) = \lambda \exp\left(-\lambda u\right),$$

where $\lambda = \dfrac{1}{\sigma_u}$. The SN distribution where the location is not 0 is given by

$$v \sim SN(\xi, \sigma_v^2, \alpha). \tag{A18}$$

The mean is of the random variable v is

$$E(v) = \xi + \sigma_v \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}. \tag{A19}$$

The assumption $E(v) = 0$ is fulfilled when

$$\xi = -\sigma_v \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}. \tag{A20}$$

Thus,

$$E\left(v + \sigma_v \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}\right) = 0.$$

We keep ξ for tractability. The density of v with a non-zero location is

$$f_v(v) = \frac{2}{\sigma_v} \phi\left(\frac{v-\xi}{\sigma_v}\right) \Phi\left(\alpha \frac{v-\xi}{\sigma_v}\right). \tag{A21}$$

Given that in our SF model, $\epsilon = v - \mathsf{p}u$, then $v - \xi = \epsilon + pu - \xi = \epsilon_r + \mathsf{p}u$, where $\epsilon_r = \epsilon - \xi$, the joint density of u and ϵ is

$$\begin{aligned} f(\epsilon, u) &= \underbrace{\frac{2}{\sigma_v} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2\right] \Phi\left(\alpha \frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)}_{f_v(v)} \underbrace{\lambda \exp\left(-\lambda u\right)}_{f_u(u)} \\ &= \frac{2\lambda}{\sigma_v} \frac{1}{\sqrt{2\pi}} \Phi\left(\alpha \frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) \exp\left[-\frac{1}{2}\left\{\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2 + 2\lambda u\right\}\right]. \end{aligned} \tag{A22}$$

A.2.1 Simplification: Sum of Powers of Exponents

Consider the expression in curly parentheses in (A22). Keeping in mind that that $\mathsf{p}^2 = 1$,

$$
\begin{aligned}
&\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2 + 2\lambda u = \frac{1}{\sigma_v^2}\left[\epsilon_r^2 + 2\mathsf{p}u\epsilon_r + \mathsf{p}^2u^2 + 2\lambda u\sigma_v^2\right] \\
&= \frac{1}{\sigma_v^2}\left[u^2 + 2u(\mathsf{p}\epsilon_r + \lambda\sigma_v^2) + \epsilon_r^2\right] \\
&= \frac{1}{\sigma_v^2}\left[u^2 + 2u(\mathsf{p}\epsilon_r + \lambda\sigma_v^2) + \epsilon_r^2 + (\mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2 - (\mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2\right] \\
&= \frac{1}{\sigma_v^2}\left[(u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2 + \epsilon_r^2 - (\mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2\right] \\
&= \frac{1}{\sigma_v^2}\left[(u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2 + \epsilon_r^2 - \mathsf{p}^2\epsilon_r^2 - 2\mathsf{p}\epsilon_r\lambda\sigma_v^2 - \lambda^2\sigma_v^4\right] \\
&= \frac{1}{\sigma_v^2}\left[(u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2 - \sigma_v^2\left(2\mathsf{p}\epsilon_r\lambda + \lambda^2\sigma_v^2\right)\right] \\
&= \frac{1}{\sigma_v^2}\left[(u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2\right] - \left(2\mathsf{p}\epsilon_r\lambda + \lambda^2\sigma_v^2\right).
\end{aligned}
\tag{A23}
$$

A.2.2 Joint Density of u and ϵ Rewritten

Using (A23), the expression in (A22) becomes

$$
\begin{aligned}
&f(\epsilon, u) = \\
&= \frac{2\lambda}{\sigma_v}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\{\left(\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)^2 + 2\lambda u\right\}\right]\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) \\
&= \frac{2\lambda}{\sigma_v}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_v^2}\left[(u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2)^2\right] - \left(2\mathsf{p}\epsilon_r\lambda + \lambda^2\sigma_v^2\right)\right\}\right]\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) \\
&= \frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)^2\right]\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) \\
&= \frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right).
\end{aligned}
\tag{A24}
$$

A.2.3 Marginal Density and Log-Likelihood

The marginal density of ϵ is obtained by integrating u out of (A24),

$$
\begin{aligned}
f(\epsilon) &= \int_0^\infty f(\epsilon, u) du \\
&= \int_0^\infty \frac{2\lambda}{\sigma_v} \exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) du \\
&= \frac{2\lambda}{\sigma_v} \exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\int_0^\infty \phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) du.
\end{aligned}
$$

To integrate

$$
\int_0^\infty \phi\left(\frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) du, \tag{A25}
$$

denote the rescaled and shifted u as

$$
u_{rs} = \frac{u + \mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v}. \tag{A26}
$$

Then $u = u_{rs}\sigma_v - \mathsf{p}\epsilon_r - \lambda\sigma_v^2$, $du = du_{rs}\sigma_v$ and (keeping in mind that $\mathsf{p}^2 = 1$)

$$
\begin{aligned}
\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v} &= \alpha\frac{\epsilon_r + \mathsf{p}(u_{rs}\sigma_v - \mathsf{p}\epsilon_r - \lambda\sigma_v^2)}{\sigma_v} \\
&= \alpha\frac{\epsilon_r + \mathsf{p}u_{rs}\sigma_v - \mathsf{p}^2\epsilon_r - \mathsf{p}\lambda\sigma_v^2}{\sigma_v} \\
&= \underbrace{-\alpha\mathsf{p}\lambda\sigma_v}_{a} + \underbrace{\alpha\mathsf{p}}_{b}\, u_{rs} \\
&= a + bu_{rs}.
\end{aligned} \tag{A27}
$$

Denote

$$
u_1 = \frac{\mathsf{p}\epsilon_r + \lambda\sigma_v^2}{\sigma_v},
$$

which is $u_{rs}|\,\{u = 0\}$. Then the integral in (A25) can be written as

$$
\sigma_v \int_{u_1}^\infty \phi\left(u_{rs}\right)\Phi\left(a + bu_{rs}\right) du_{rs} \tag{A28}
$$

and

$$f(\epsilon) = 2\lambda \exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\int_{u_1}^{\infty}\phi\left(u_{rs}\right)\Phi\left(a + bu_{rs}\right)du_{rs}. \tag{A29}$$

Using the result in (A11)[21]

$$\begin{aligned}
\mathcal{A} &= \int_{u_1}^{\infty}\phi\left(u_{rs}\right)\Phi\left(a+bu_{rs}\right)du_{rs} \\
&= -T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right) \\
&\quad + T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right)\Phi\left(-u_1\right).
\end{aligned} \tag{A30}$$

Then,

$$\begin{aligned}
f(\epsilon) &= 2\lambda \exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right) \times \mathcal{A} \\
&= 2\lambda \exp\left(\mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2}\right)\left[\begin{array}{l} -T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right) \\ +T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right)\Phi\left(-u_1\right) \end{array}\right],
\end{aligned} \tag{A31}$$

where $a = -\alpha\mathsf{p}\lambda\sigma_v$, $b = \alpha\mathsf{p}$, $a_2 = a/\sqrt{1+b^2}$, and $u_1 = \mathsf{p}\epsilon_r/\sigma_v + \lambda\sigma_v$. The log-likelihood based on (A31) is therefore

$$logl = \ln\left(2\lambda\right) + \mathsf{p}\epsilon_r\lambda + \frac{\lambda^2\sigma_v^2}{2} + \ln\mathcal{A}. \tag{A32}$$

A.2.4 Gradient

To derive the gradient, note that

$$\frac{\partial T(h,a)}{\partial\omega} = T_1'(h,a)\frac{\partial h}{\partial\omega} + T_2'(h,a)\frac{\partial a}{\partial\omega}, \tag{A33}$$

[21] The result (A16) can also be used.

where

$$T_1'(h,a) = \frac{\partial T(h,a)}{\partial h} = -\frac{1}{2\sqrt{2\pi}}(2\Phi(ha)-1)\exp\left(-\frac{h^2}{2}\right) = -\phi(h)(\Phi(ha)-0.5)$$

and

$$T_2'(h,a) = \frac{\partial T(h,a)}{\partial a} = \frac{1}{2\pi(1+a^2)}\exp\left(-\frac{h^2}{2}\left(1+a^2\right)\right).$$

Further denote,

$$A = \alpha/\sqrt{1+\alpha^2}$$
$$SVU = \lambda\sigma_v.$$

Given our specifications, the following notations for derivatives are needed

$$\epsilon_\beta = -\boldsymbol{x}$$
$$\alpha_s = -\boldsymbol{z}_s$$
$$A_s = \frac{\partial A}{\partial \boldsymbol{\gamma}_s} = (1.0+\alpha^2)^{-3/2}\boldsymbol{z}_s$$
$$SV1 = \frac{\partial(1/\sigma_v)}{\partial \boldsymbol{\gamma}_v} = -0.5\frac{\boldsymbol{z}_v}{\sigma_v}$$
$$SU1 = \frac{\partial(1/\sigma_u)}{\partial \boldsymbol{\gamma}_u} = -0.5\frac{\boldsymbol{z}_u}{\sigma_u}$$
$$SVU_u = -0.5SVU\boldsymbol{z}_u$$
$$SVU_v = 0.5SVU\boldsymbol{z}_v.$$

The partial derivatives of u_1 are:

$$\frac{\partial u_1}{\partial \boldsymbol{\beta}} = \frac{\mathsf{p}\epsilon_\beta}{\sigma_v}$$
$$\frac{\partial u_1}{\partial \boldsymbol{\gamma}_v} = \mathsf{p}\epsilon SV1 + SVU_v$$
$$\frac{\partial u_1}{\partial \boldsymbol{\gamma}_s} = \sqrt{\frac{2}{\pi}}A_s$$

$$\frac{\partial u_1}{\partial \boldsymbol{\gamma}_u} = SVU_u.$$

The partial derivatives of a_2 are:

$$\begin{aligned}
\frac{\partial a_2}{\partial \boldsymbol{\beta}} &= \mathbf{0} \\
\frac{\partial a_2}{\partial \boldsymbol{\gamma}_v} &= -\mathsf{p}A \times SVU_v \\
\frac{\partial a_2}{\partial \boldsymbol{\gamma}_s} &= -\mathsf{p}A_s \times SVU \\
\frac{\partial a_2}{\partial \boldsymbol{\gamma}_u} &= -\mathsf{p}A \times SVU_u.
\end{aligned}$$

The partial derivatives of a_2/u_1 are:

$$\begin{aligned}
\frac{\partial (a_2/u_1)}{\partial \boldsymbol{\beta}} &= \frac{-a_2 \frac{\partial u_1}{\partial \boldsymbol{\beta}}}{u_1^2} \\
\frac{\partial (a_2/u_1)}{\partial \boldsymbol{\gamma}_v} &= \frac{u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_v} - a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_v}}{u_1^2} \\
\frac{\partial (a_2/u_1)}{\partial \boldsymbol{\gamma}_s} &= \frac{u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_s} - a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_s}}{u_1^2} \\
\frac{\partial (a_2/u_1)}{\partial \boldsymbol{\gamma}_u} &= \frac{u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_u} - a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_u}}{u_1^2}.
\end{aligned}$$

The partial derivatives of u_1/a_2 are:

$$\begin{aligned}
\frac{\partial (u_1/a_2)}{\partial \boldsymbol{\beta}} &= \frac{-a_2 \frac{\partial u_1}{\partial \boldsymbol{\beta}}}{a_2^2} = \frac{-\frac{\partial u_1}{\partial \boldsymbol{\beta}}}{a_2} \\
\frac{\partial (u_1/a_2)}{\partial \boldsymbol{\gamma}_v} &= \frac{a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_v} - u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_v}}{a_2^2} \\
\frac{\partial (u_1/a_2)}{\partial \boldsymbol{\gamma}_s} &= \frac{a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_s} - u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_s}}{a_2^2} \\
\frac{\partial (u_1/a_2)}{\partial \boldsymbol{\gamma}_u} &= \frac{a_2 \frac{\partial u_1}{\partial \boldsymbol{\gamma}_u} - u_1 \frac{\partial a_2}{\partial \boldsymbol{\gamma}_u}}{a_2^2}.
\end{aligned}$$

The partial derivatives of a are:

$$
\begin{aligned}
\frac{\partial a}{\partial \boldsymbol{\beta}} &= \mathbf{0} \\
\frac{\partial a}{\partial \boldsymbol{\gamma}_v} &= -\mathsf{p}\alpha \times SVU_v \\
\frac{\partial a}{\partial \boldsymbol{\gamma}_s} &= -\mathsf{p}\alpha_s \times SVU \\
\frac{\partial a}{\partial \boldsymbol{\gamma}_u} &= -\mathsf{p}\alpha \times SVU_u.
\end{aligned}
$$

The partial derivatives of $b + a/u_1$ are:

$$
\begin{aligned}
\frac{\partial(b + a/u_1)}{\partial \boldsymbol{\beta}} &= \frac{-a\frac{\partial u_1}{\partial \boldsymbol{\beta}}}{u_1^2} \\
\frac{\partial(b + a/u_1)}{\partial \boldsymbol{\gamma}_v} &= \frac{u_1\frac{\partial a}{\partial \boldsymbol{\gamma}_v} - a\frac{\partial u_1}{\partial \boldsymbol{\gamma}_v}}{u_1^2} \\
\frac{\partial(b + a/u_1)}{\partial \boldsymbol{\gamma}_s} &= \frac{u_1\frac{\partial a}{\partial \boldsymbol{\gamma}_s} - a\frac{\partial u_1}{\partial \boldsymbol{\gamma}_s}}{u_1^2} + \mathsf{p}\alpha_s \\
\frac{\partial(b + a/u_1)}{\partial \boldsymbol{\gamma}_u} &= \frac{u_1\frac{\partial a}{\partial \boldsymbol{\gamma}_u} - a\frac{\partial u_1}{\partial \boldsymbol{\gamma}_u}}{u_1^2}.
\end{aligned}
$$

The partial derivatives of $b + u_1(1 + b^2)/a$ are:

$$
\begin{aligned}
\frac{\partial(b + u_1(1 + b^2)/a)}{\partial \boldsymbol{\beta}} &= \frac{(1 + b^2)\frac{\partial u_1}{\partial \boldsymbol{\beta}}}{a} \\
\frac{\partial(b + u_1(1 + b^2)/a)}{\partial \boldsymbol{\gamma}_v} &= (1 + b^2)\frac{a\frac{\partial u_1}{\partial \boldsymbol{\gamma}_v} - u_1\frac{\partial a}{\partial \boldsymbol{\gamma}_v}}{a^2} \\
\frac{\partial(b + u_1(1 + b^2)/a)}{\partial \boldsymbol{\gamma}_s} &= \frac{a(1 + b^2)\frac{\partial u_1}{\partial \boldsymbol{\gamma}_s} - u1/SVU^2(1/b^2 - 1)\frac{\partial a}{\partial \boldsymbol{\gamma}_s}}{a^2} + \mathsf{p}\alpha_s \\
\frac{\partial(b + u_1(1 + b^2)/a)}{\partial \boldsymbol{\gamma}_u} &= (1 + b^2)\frac{a\frac{\partial u_1}{\partial \boldsymbol{\gamma}_u} - u_1\frac{\partial a}{\partial \boldsymbol{\gamma}_u}}{a^2}.
\end{aligned}
$$

The partial derivatives of the first three terms (denoted by $logl_{-\ln \mathcal{A}}$) in (A32) are:

$$
\frac{\partial logl_{-\ln \mathcal{A}}}{\partial \boldsymbol{\beta}} = \frac{\mathsf{p}\epsilon_{\beta}}{\sigma_u}
$$

$$\frac{\partial logl_{-\ln\mathcal{A}}}{\partial\boldsymbol{\gamma}_v} = \left(\mathsf{p}\sqrt{\frac{2}{\pi}}A + SVU\right) SVU_v$$
$$\frac{\partial logl_{-\ln\mathcal{A}}}{\partial\boldsymbol{\gamma}_s} = \sqrt{\frac{2}{\pi}}A_s$$
$$\frac{\partial logl_{-\ln\mathcal{A}}}{\partial\boldsymbol{\gamma}_u} = \mathsf{p}\epsilon SU1 + \left(\mathsf{p}\sqrt{\frac{2}{\pi}}A + SVU\right) SVU_u - 0.5\boldsymbol{z}_u.$$

Note that

$$\begin{aligned}\frac{\partial[\Phi(a_2)\Phi(-u_1)]}{\partial\omega} &= \Phi(a_2)\frac{\partial\Phi(-u_1)}{\partial\omega} + \Phi(-u_1)\frac{\partial\Phi(a_2)}{\partial\omega}\\ &= -\Phi(a_2)\phi(-u_1)\frac{\partial u_1}{\partial\omega} + \Phi(-u_1)\phi(a_2)\frac{\partial a_2}{\partial\omega}.\end{aligned}$$

Denoting $\Phi_1 = -\Phi(a_2)\phi(-u_1)$ and $\Phi_2 = \Phi(-u_1)\phi(a_2)$, the partial derivatives of $\Phi(a_2)\Phi(-u_1)$ are:

$$\frac{\partial(\Phi(a_2)\Phi(-u_1))}{\partial\boldsymbol{\beta}} = \Phi_1\frac{\partial u_1}{\partial\boldsymbol{\beta}} + \Phi_2\frac{\partial a_2}{\partial\boldsymbol{\beta}}$$
$$\frac{\partial(\Phi(a_2)\Phi(-u_1))}{\partial\boldsymbol{\gamma}_v} = \Phi_1\frac{\partial u_1}{\partial\boldsymbol{\gamma}_v} + \Phi_2\frac{\partial a_2}{\partial\boldsymbol{\gamma}_v}$$
$$\frac{\partial(\Phi(a_2)\Phi(-u_1))}{\partial\boldsymbol{\gamma}_s} = \Phi_1\frac{\partial u_1}{\partial\boldsymbol{\gamma}_s} + \Phi_2\frac{\partial a_2}{\partial\boldsymbol{\gamma}_s}$$
$$\frac{\partial(\Phi(a_2)\Phi(-u_1))}{\partial\boldsymbol{\gamma}_u} = \Phi_1\frac{\partial u_1}{\partial\boldsymbol{\gamma}_u} + \Phi_2\frac{\partial a_2}{\partial\boldsymbol{\gamma}_u}.$$

Now we can collect the terms recalling (A33) to obtain the gradient in the direction of $\boldsymbol{\beta}$:

$$\frac{\partial logl}{\partial\boldsymbol{\beta}} = \frac{\partial logl_{-\ln\mathcal{A}}}{\partial\boldsymbol{\beta}}$$

$$
+\frac{1}{\mathcal{A}}\left[\begin{array}{l}
-\left[T_1'\left(u_1,\frac{a_2}{u_1}\right)\frac{\partial u_1}{\partial\boldsymbol{\beta}}+T_2'\left(u_1,\frac{a_2}{u_1}\right)\frac{\partial\left(a_2/u_1\right)}{\partial\boldsymbol{\beta}}\right]\\
-\left[T_1'\left(a_2,\frac{u_1}{a_2}\right)\frac{\partial a_2}{\partial\boldsymbol{\beta}}+T_2'\left(a_2,\frac{u_1}{a_2}\right)\frac{\partial\left(u_1/a_2\right)}{\partial\boldsymbol{\beta}}\right]\\
+\left[T_1'\left(u_1,b+\frac{a}{u_1}\right)\frac{\partial u_1}{\partial\boldsymbol{\beta}}+T_2'\left(u_1,b+\frac{a}{u_1}\right)\frac{\partial\left(b+\frac{a}{u_1}\right)}{\partial\boldsymbol{\beta}}\right]\\
+\left[T_1'\left(a_2,b+\frac{u_1(1+b^2)}{a}\right)\frac{\partial a_2}{\partial\boldsymbol{\beta}}\right.\\
\qquad\left.+T_2'\left(a_2,b+\frac{u_1(1+b^2)}{a}\right)\frac{\partial\left(b+\frac{u_1(1+b^2)}{a}\right)}{\partial\boldsymbol{\beta}}\right]\\
+\left[\Phi_1\frac{\partial u_1}{\partial\boldsymbol{\beta}}+\Phi_2\frac{\partial a_2}{\partial\boldsymbol{\beta}}\right]
\end{array}\right]. \tag{A34}
$$

The gradient in the direction of $\boldsymbol{\gamma}_v$, $\boldsymbol{\gamma}_s$, and $\boldsymbol{\gamma}_u$ is obtained by replacing $\boldsymbol{\beta}$ with $\boldsymbol{\gamma}_v$, $\boldsymbol{\gamma}_s$, and $\boldsymbol{\gamma}_u$ in (A34).

A.3 Efficiency Estimation

To obtain observation-specific estimates of inefficiency u, we follow Jondrow et al. (1982) and first obtain the conditional distribution of u given ϵ.

$$
\begin{aligned}
f(u|\epsilon) &= \frac{f(u,\epsilon)}{f(\epsilon)}\\
&= \frac{\frac{2\lambda}{\sigma_v}\exp\left(\mathsf{p}\epsilon_r\lambda+\frac{\lambda^2\sigma_v^2}{2}\right)\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{2\lambda\exp\left(\mathsf{p}\epsilon_r\lambda+\frac{\lambda^2\sigma_v^2}{2}\right)\mathcal{A}}\\
&= \frac{\frac{1}{\sigma_v}\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{\mathcal{A}}.
\end{aligned} \tag{A35}
$$

We then obtain the point estimator for u by finding the mean value of the conditional distribution in (A35)

$$
E(u|\epsilon) = \int_0^{+\infty}\frac{u\times\frac{1}{\sigma_v}\phi\left(\frac{u+\mathsf{p}\epsilon_r+\lambda\sigma_v^2}{\sigma_v}\right)\Phi\left(\alpha\frac{\epsilon_r+\mathsf{p}u}{\sigma_v}\right)}{\mathcal{A}}du. \tag{A36}
$$

Using the earlier notation on a, b, u_1, denoting $u/\sigma_v + u_1 = u_{rs}$, which implies $du/\sigma_v = du_{rs}$ and noting that $\mathcal{A}$ does not depend on u, the integral in (A36) can be rewritten as

$$E(u|\epsilon) = \frac{1}{\mathcal{A}} \int_0^{+\infty} u \times \frac{1}{\sigma_v} \phi\left(\frac{u}{\sigma_v} + u_1\right) \Phi\left(\alpha \frac{\epsilon_r + \mathsf{p} u}{\sigma_v}\right) du \tag{A37}$$

$$= \frac{1}{\mathcal{A}} \int_{u_1}^{+\infty} (u_{rs}\sigma_v - u_1\sigma_v) \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs}. \tag{A38}$$

The integral in (A38) is split into 2 parts

$$\begin{aligned}
E(u|\epsilon) &= \frac{1}{\mathcal{A}} \left(\int_{u_1}^{+\infty} -u_1\sigma_v \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right. \\
&\qquad \left. + \int_{u_1}^{+\infty} u_{rs}\sigma_v \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right) \\
&= \frac{1}{\mathcal{A}} \left(-u_1\sigma_v \times \int_{u_1}^{+\infty} \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right. \\
&\qquad \left. + \sigma_v \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right) \\
&= \frac{1}{\mathcal{A}} \left(-u_1\sigma_v \times \mathcal{A} \right. \\
&\qquad \left. + \sigma_v \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right) \\
&= -u_1\sigma_v + \frac{\sigma_v}{\mathcal{A}} \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \\
&= \sigma_v \left(\left(\frac{1}{\mathcal{A}} \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs} \right) - u_1 \right).
\end{aligned} \tag{A39}$$

Using the result in Owen (1980),

$$\begin{aligned}
\int \boldsymbol{x}\phi(\boldsymbol{x}) \Phi(\boldsymbol{a} + \boldsymbol{b}\boldsymbol{x}) d\boldsymbol{x} &= \frac{\boldsymbol{b}}{\sqrt{1+\boldsymbol{b}^2}} \phi\left(\frac{\boldsymbol{a}}{\sqrt{1+\boldsymbol{b}^2}}\right) \Phi\left(\boldsymbol{x}\sqrt{1+\boldsymbol{b}^2} + \frac{\boldsymbol{a}\boldsymbol{b}}{\sqrt{1+\boldsymbol{b}^2}}\right) \\
&\quad - \phi(\boldsymbol{x}) \Phi(\boldsymbol{a} + \boldsymbol{b}\boldsymbol{x}),
\end{aligned} \tag{A40}$$

the integral in the first part of equation (A39) can be written as

$$\int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs}) \Phi(a + bu_{rs}) du_{rs}$$

$$= \left[\frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \Phi\left(u_{rs}\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}\right) - \phi\left(u_{rs}\right) \Phi\left(a + bu_{rs}\right) \right] \Bigg|_{u_{rs}=u_1}^{u_{rs}=+\infty}$$

$$= \frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \left\{ 1 - \Phi\left(u_1\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}\right) \right\}$$

$$- \left\{ 0 - \phi\left(u_1\right) \Phi\left(a + bu_1\right) \right\}. \tag{A41}$$

Therefore,

$$E(u|\epsilon) = \tag{A42}$$

$$= \sigma_v \left(\frac{\text{expression in } (A41)}{\mathcal{A}} - u_1 \right)$$

$$= \sigma_v \left(\frac{\left[\begin{array}{l} \frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \Phi\left(-u_1\sqrt{1+b^2} - \frac{ab}{\sqrt{1+b^2}}\right) + \\ \phi\left(u_1\right) \Phi\left(a + bu_1\right) \end{array} \right]}{\left[\begin{array}{l} -T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right) \\ +T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right) \Phi\left(-u_1\right) \end{array} \right]} - u_1 \right)$$

$$= -\mathsf{p}\epsilon_r - \lambda\sigma_v^2 + \sigma_v \frac{\left[\begin{array}{l} \frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \Phi\left(-u_1\sqrt{1+b^2} - \frac{ab}{\sqrt{1+b^2}}\right) + \\ \phi\left(u_1\right) \Phi\left(a + bu_1\right) \end{array} \right]}{\left[\begin{array}{l} -T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right) \\ +T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right) \Phi\left(-u_1\right) \end{array} \right]}.$$

B Skew-Normal Truncated Normal Stochastic Frontier Model

Consider the model is described by (A17), but now assume that u is truncated normally distributed on a non-negative space,

$$u \sim TN(\mu, \sigma_u^2, 0, \infty). \tag{B1}$$

B.1 Derivation of $f(\epsilon)$ and the Log-Likelihood Function

It still holds that $\epsilon = v - \mathsf{p}u$, then $v - \xi = \epsilon + pu - \xi$, where ξ is defined in (A20). The joint density of u and ϵ can be written as

$$\begin{aligned} f(\epsilon, u) = &\Phi\left(\alpha \frac{\epsilon + \mathsf{p}u - \xi}{\sigma_v}\right) \frac{1}{\Phi(\mu/\sigma_u)} \\ &\times \frac{2}{\sigma_v \sigma_u} \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left\{\left(\frac{\epsilon + \mathsf{p}u - \xi}{\sigma_v}\right)^2 + \left(\frac{u-\mu}{\sigma_u}\right)^2\right\}\right). \end{aligned} \tag{B2}$$

B.1.1 Simplification: Sum of Powers of Exponents

First, consider the sum of exponents in (B2):

$$\left(\frac{\epsilon + \mathsf{p}u - \xi}{\sigma_v}\right)^2 + \left(\frac{u-\mu}{\sigma_u}\right)^2 = \frac{1}{\sigma_u^2 \sigma_u^2} \begin{bmatrix} \sigma_u^2(\epsilon^2 + \mathsf{p}^2 u^2 + \xi^2 + \\ 2\epsilon \mathsf{p} u - 2\epsilon\xi - 2pu\xi) + \\ \sigma_v^2(u^2 - 2u\mu + \mu^2) \end{bmatrix}. \tag{B3}$$

Now consider only the expression in squared brackets, where we need to complete the square

$$\begin{aligned} &\epsilon^2\sigma_u^2 + u^2\sigma_u^2 + \xi^2\sigma_u^2 + 2\epsilon\mathsf{p}u\sigma_u^2 - 2\epsilon\xi\sigma_u^2 - 2\mathsf{p}u\xi\sigma_u^2 + u^2\sigma_v^2 - 2u\mu\sigma_v^2 + \mu^2\sigma_v^2 + \\ &u^2(\sigma_u^2 + \sigma_v^2) + 2u(\epsilon\mathsf{p}\sigma_u^2 - \mathsf{p}\xi\sigma_u^2 - \mu\sigma_v^2) + \sigma_u^2(\epsilon^2 + \xi^2 - 2\epsilon\xi) + \mu^2\sigma_v^2. \end{aligned} \tag{B4}$$

Denote relocated ϵ by $\epsilon_r = \epsilon - \xi$ and rcall that $\mathsf{p}^2 = 1$, then (B4) can be written as

$$u^2(\sigma_u^2 + \sigma_v^2) + 2u(\mathsf{p}\sigma_u^2(\epsilon - \xi) - \mu\sigma_v^2) + (\epsilon - \xi)^2\sigma_u^2 + \mu^2\sigma_v^2$$

or

$$u^2(\sigma_u^2 + \sigma_v^2) + 2u(\epsilon_r\mathsf{p}\sigma_u^2 - \mu\sigma_v^2) + \epsilon_r^2\sigma_u^2 + \mu^2\sigma_v^2. \tag{B5}$$

B.1.2 Joint Density of u and ϵ

We can now rewrite the expression in (B2)

$$f(\epsilon, u) = \left[\frac{2}{\Phi(\mu/\sigma_u)}\frac{1}{\sigma\sigma_*}\phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right)\right]\phi\left(\frac{u-\mu_{1,r}}{\sigma_*}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right), \tag{B6}$$

where

$$\mu_{1,r} = \frac{\mu\sigma_v^2 - \epsilon_r\mathsf{p}\sigma_u^2}{\sigma^2}, \tag{B7}$$

and

$$\epsilon_r = \epsilon + \sigma_v\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}} = y - f(\boldsymbol{x};\boldsymbol{\beta}) + \sigma_v\sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}}. \tag{B8}$$

B.1.3 Marginal Density and Log-Likelihood

The marginal density of ϵ is obtained by integrating u out (B6) over the domain of u

$$\int_0^\infty f(\epsilon, u)du = \frac{2}{\Phi(\mu/\sigma_u)}\frac{1}{\sigma\sigma_*}\phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right)\int_0^\infty \phi\left(\frac{u-\mu_{1,r}}{\sigma_*}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)du. \tag{B9}$$

Denote the relocated and scaled variant of u

$$u_{rs} = \frac{u-\mu_{1,r}}{\sigma_*}. \tag{B10}$$

Then $u = \mu_{1,r} + u_{rs}\sigma_*$ and

$$\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v} = \alpha\frac{\epsilon_r + \mathsf{p}(\mu_{1,r} + u_{rs}\sigma_*)}{\sigma_v} = \alpha\frac{\epsilon_r + \mathsf{p}\mu_{1,r}}{\sigma_v} + \alpha\frac{\mathsf{p}\sigma_*}{\sigma_v}u_{rs} = a + bu_{rs}. \tag{B11}$$

Then the marginal density of ϵ given in (B9) is

$$f(\epsilon) = \frac{2}{\Phi(\mu/\sigma_u)}\frac{1}{\sigma\sigma_*}\phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right)\int_{-\frac{\mu_1}{\sigma_*}}^\infty \phi\left(u_{rs}\right)\Phi\left(a + bu_{rs}\right)\sigma_* du_{rs} \tag{B12}$$

$$= \frac{2}{\Phi(\mu/\sigma_u)}\frac{1}{\sigma}\phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right)\int_{-\frac{\mu_1}{\sigma_*}}^\infty \phi\left(u_{rs}\right)\Phi\left(a + bu_{rs}\right)du_{rs}. \tag{B13}$$

Denoting $u_1 = -\mu_1/\sigma_*$ and using the result in (A11)[22]

$$\begin{aligned} \mathcal{B} &= \int_{u_1}^{\infty} \phi\left(u_{rs}\right) \Phi\left(a + b u_{rs}\right) du_{rs} \\ &= -T\left(u_1, \frac{a_2}{u_1}\right) - T\left(a_2, \frac{u_1}{a_2}\right) + T\left(u_1, b + \frac{a}{u_1}\right) \\ &\quad + T\left(a_2, b + \frac{u_1(1+b^2)}{a}\right) + \Phi\left(a_2\right) \Phi\left(-u_1\right), \end{aligned} \tag{B14}$$

where $a = \alpha(\epsilon_r + \mathsf{p}\mu_{1,r})/\sigma_v$ and $b = \alpha \mathsf{p} \sigma_*/\sigma_v$. Then

$$f(\epsilon) = \frac{2}{\Phi(\mu/\sigma_u)} \frac{1}{\sigma} \phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right) \times \mathcal{B}. \tag{B15}$$

The log-likelihood based on (B15) is therefore

$$\ln(2) - \ln \Phi\left(\frac{\mu}{\sigma_u}\right) - \ln \sigma + \ln \phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right) + \ln \mathcal{B}. \tag{B16}$$

The gradients in case of the truncated normal u are derived analogously to the case of exponential u, shown in A.2.4.

B.2 Efficiency Estimation

To obtain observation-specific estimates of inefficiency u, we follow Jondrow et al. (1982) and first obtain the conditional distribution of u given ϵ

$$\begin{aligned} f(u|\epsilon) &= \frac{f(\epsilon, x)}{f(\epsilon)} \\ &= \frac{\left[\frac{2}{\Phi(\mu/\sigma_u)} \frac{1}{\sigma\sigma_*} \phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right)\right] \phi\left(\frac{u - \mu_{1,r}}{\sigma_*}\right) \Phi\left(\alpha \frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)}{\frac{2}{\Phi(\mu/\sigma_u)} \frac{1}{\sigma} \phi\left(\frac{\mu + \mathsf{p}\epsilon_r}{\sigma}\right) \times \mathcal{B}} \\ &= \frac{\frac{1}{\sigma_*} \phi\left(\frac{u - \mu_{1,r}}{\sigma_*}\right) \Phi\left(\alpha \frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)}{\mathcal{B}.} \end{aligned} \tag{B17}$$

[22] The result (A16) can also be used.

We then obtain the point estimator for u by finding the mean value of the conditional distribution in (B17)

$$E(u|\epsilon) = \int_0^{+\infty} \frac{u \times \frac{1}{\sigma_*}\phi\left(\frac{u-\mu_{1,r}}{\sigma_*}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right)}{\mathcal{B}} du. \tag{B18}$$

Since $\mathcal{B}$ do not depend on u,

$$E(u|\epsilon) = \frac{1}{\mathcal{B}\sigma_*}\int_0^{+\infty} u \times \phi\left(\frac{u-\mu_{1,r}}{\sigma_*}\right)\Phi\left(\alpha\frac{\epsilon_r + \mathsf{p}u}{\sigma_v}\right) du. \tag{B19}$$

To make use of the (A40), recall the notation (B10) and (B11). Then the integral in (B19) becomes

$$\begin{aligned} E(u|\epsilon) &= \frac{1}{\mathcal{B}\sigma_*}\int_{u_1}^{+\infty} (\mu_{1,r} + u_{rs}\sigma_*) \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,\sigma_* du_{rs} \\ &= \frac{1}{\mathcal{B}}\int_{u_1}^{+\infty} (\mu_{1,r} + u_{rs}\sigma_*) \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}, \end{aligned} \tag{B20}$$

where $u_1 = -\mu_{1,r}/\sigma_*$, as before by setting $u = 0$ into (B10). The integral in (B20) is split into 2 parts

$$\begin{aligned} E(u|\epsilon) &= \frac{1}{\mathcal{B}}\left(\int_{u_1}^{+\infty} \mu_{1,r} \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}\right. \\ &\qquad \left. + \int_{u_1}^{+\infty} u_{rs}\sigma_* \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}\right) \\ &= \frac{1}{\mathcal{B}}\left(\mu_{1,r} \times \int_{u_1}^{+\infty} \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}\right. \\ &\qquad \left. + \sigma_* \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}\right) \\ &= \frac{1}{\mathcal{B}}\left(\mu_{1,r} \times \mathcal{B}\right. \\ &\qquad \left. + \sigma_* \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}\right) \\ &= \mu_{1,r} + \frac{\sigma_*}{\mathcal{B}} \times \int_{u_1}^{+\infty} u_{rs} \times \phi(u_{rs})\,\Phi(a + bu_{rs})\,du_{rs}. \end{aligned} \tag{B21}$$

Using (A40),

$$
\begin{aligned}
&\int_{u_1}^{+\infty} u_{rs} \times \phi\left(u_{rs}\right) \Phi\left(a+b u_{rs}\right) d u_{rs} \\
&\quad=\left[\frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \Phi\left(u_{rs} \sqrt{1+b^2}+\frac{a b}{\sqrt{1+b^2}}\right)-\phi\left(u_{rs}\right) \Phi\left(a+b u_{rs}\right)\right]\Bigg|_{u_{rs}=u_1}^{u_{rs}=+\infty} \\
&\quad=\frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right)\left\{1-\Phi\left(u_1 \sqrt{1+b^2}+\frac{a b}{\sqrt{1+b^2}}\right)\right\} \\
&\quad-\left\{0-\phi\left(u_1\right) \Phi\left(a+b u_1\right)\right\} .
\end{aligned} \tag{B22}
$$

Therefore,

$$
\begin{aligned}
E(u|\epsilon) &= \mu_{1,r} + \frac{\sigma_*}{\mathcal{B}} \times (\text{expression in } (B22)) \\
&= \mu_{1,r} + \sigma_* \frac{\left[\begin{array}{l} \frac{b}{\sqrt{1+b^2}} \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \Phi\left(-u_1 \sqrt{1+b^2}-\frac{a b}{\sqrt{1+b^2}}\right)+ \\ \phi\left(u_1\right) \Phi\left(a+b u_1\right) \end{array}\right]}{\left[\begin{array}{l} -T\left(u_1, \frac{a_2}{u_1}\right)-T\left(a_2, \frac{u_1}{a_2}\right)+T\left(u_1, \frac{a}{u_1}+b\right)+ \\ T\left(a_2, \frac{u_1 a}{a_2^2}+b\right)+\Phi\left(a_2\right) \Phi\left(-u_1\right) \end{array}\right]} .
\end{aligned} \tag{B23}
$$

C Simulation study and likelihood profile analysis

C.1 SN regression

Table C1: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$. Bias and MSE, $\sigma_v = 0.5$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 0.5$, $\alpha = 0.5$								
250	0.0011	0.0023	−0.0009	0.0029	0.0466	0.0024	−0.0607	0.6667
500	−0.0014	0.0012	0.0017	0.0014	0.0340	0.0013	0.0152	0.3450
1000	−0.0003	0.0005	0.0004	0.0007	0.0206	0.0007	−0.0409	0.1960
2000	0.0014	0.0003	−0.0007	0.0003	0.0131	0.0004	−0.0173	0.0916
$\sigma_v = 0.5$, $\alpha = 1$								
250	−0.0006	0.0018	−0.0008	0.0022	0.0008	0.0014	0.0005	0.1417
500	−0.0003	0.0009	0.0006	0.0010	−0.0003	0.0009	0.0028	0.0637
1000	0.0003	0.0004	−0.0005	0.0005	−0.0029	0.0005	−0.0110	0.0348
2000	0.0006	0.0002	−0.0006	0.0003	−0.0014	0.0003	−0.0110	0.0162
$\sigma_v = 0.5$, $\alpha = 2$								
250	−0.0006	0.0012	0.0012	0.0016	−0.0043	0.0008	−0.0044	0.1289
500	0.0001	0.0007	−0.0004	0.0007	−0.0024	0.0004	−0.0002	0.0567
1000	0.0002	0.0003	−0.0001	0.0003	−0.0004	0.0002	−0.0029	0.0274
2000	−0.0000	0.0001	−0.0004	0.0002	−0.0009	0.0001	−0.0086	0.0148

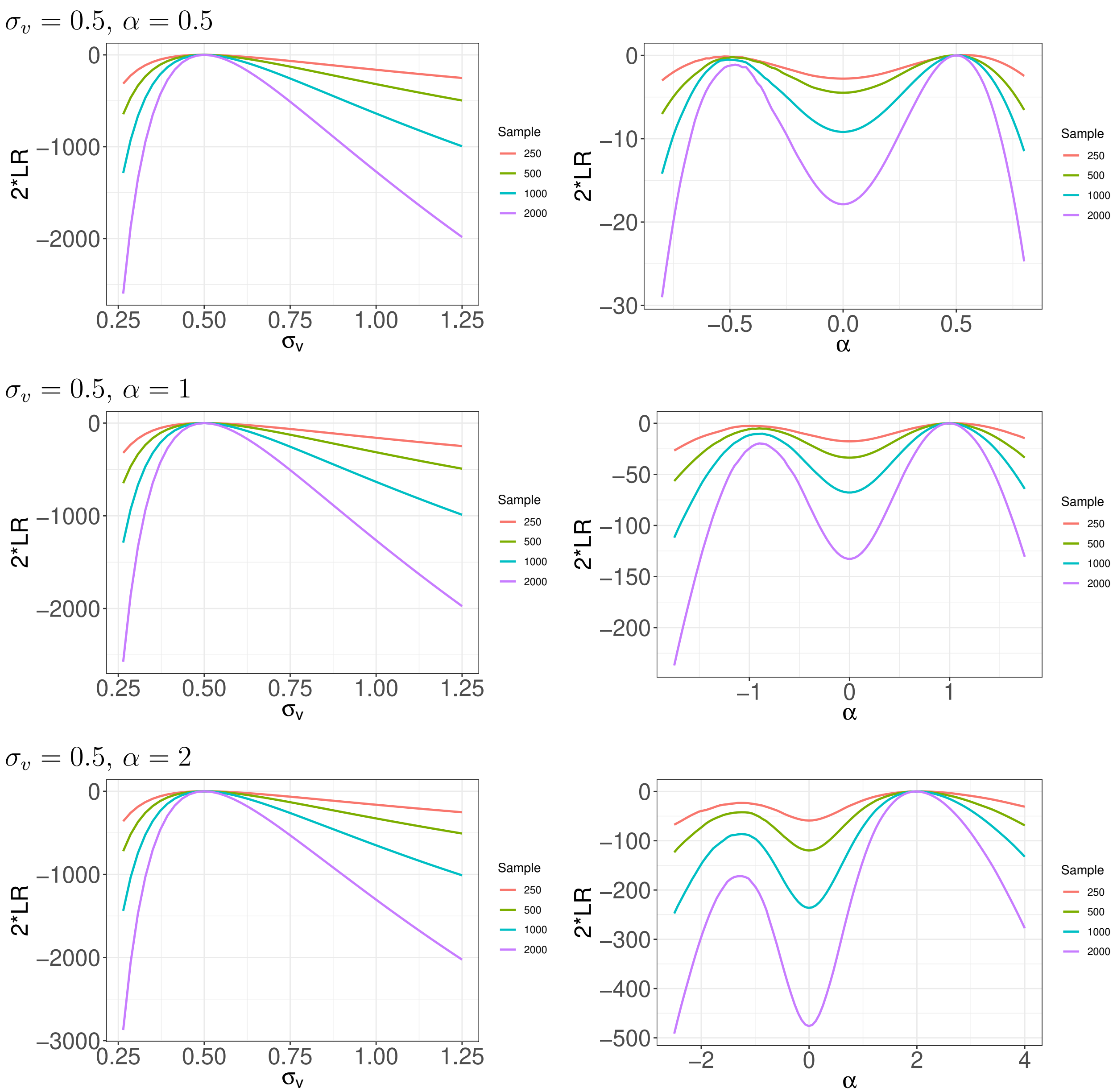
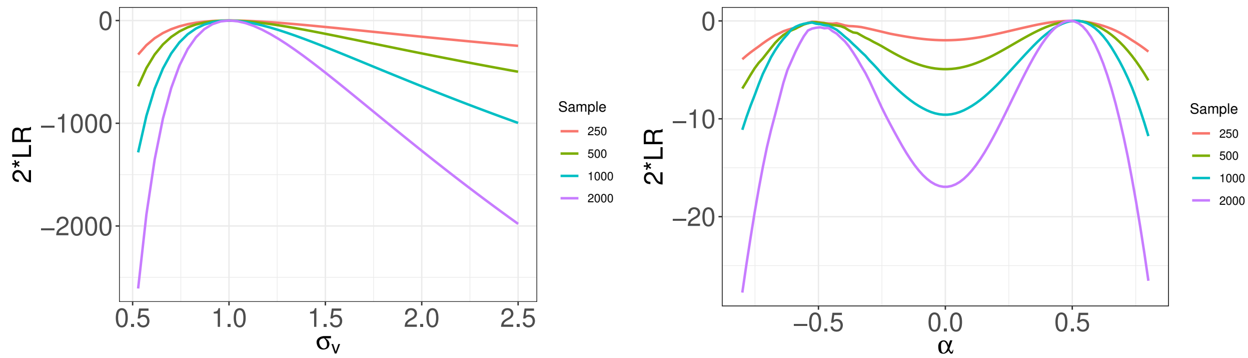


Figure C1: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

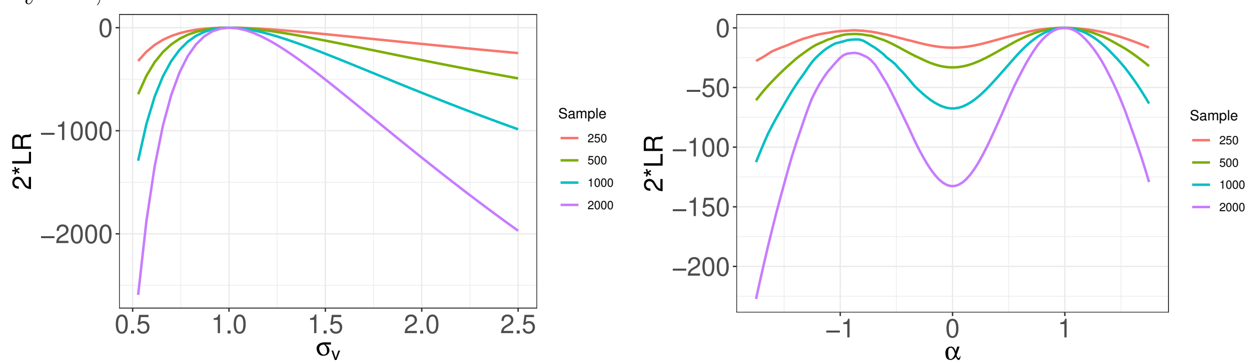
Table C2: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$. Bias and MSE, $\sigma_v = 1$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 1$, $\alpha = 0.5$								
250	0.0023	0.0093	−0.0019	0.0114	0.0930	0.0094	−0.0619	0.6683
500	−0.0027	0.0047	0.0034	0.0054	0.0681	0.0054	0.0146	0.3466
1000	−0.0006	0.0021	0.0009	0.0026	0.0411	0.0026	−0.0401	0.1961
2000	0.0029	0.0012	−0.0015	0.0014	0.0263	0.0016	−0.0176	0.0914
$\sigma_v = 1$, $\alpha = 1$								
250	−0.0012	0.0071	−0.0015	0.0090	0.0015	0.0057	0.0001	0.1417
500	−0.0007	0.0038	0.0012	0.0039	−0.0005	0.0036	0.0024	0.0639
1000	0.0005	0.0017	−0.0009	0.0021	−0.0059	0.0022	−0.0111	0.0349
2000	0.0012	0.0009	−0.0011	0.0011	−0.0029	0.0011	−0.0108	0.0162
$\sigma_v = 1$, $\alpha = 2$								
250	−0.0011	0.0049	0.0023	0.0062	−0.0087	0.0034	−0.0044	0.1295
500	0.0001	0.0027	−0.0008	0.0028	−0.0047	0.0016	−0.0000	0.0567
1000	0.0004	0.0011	−0.0003	0.0013	−0.0008	0.0007	−0.0033	0.0274
2000	−0.0001	0.0006	−0.0008	0.0007	−0.0019	0.0004	−0.0086	0.0148

$\sigma_v = 1,\ \alpha = 0.5$



$\sigma_v = 1,\ \alpha = 1$



$\sigma_v = 1,\ \alpha = 2$

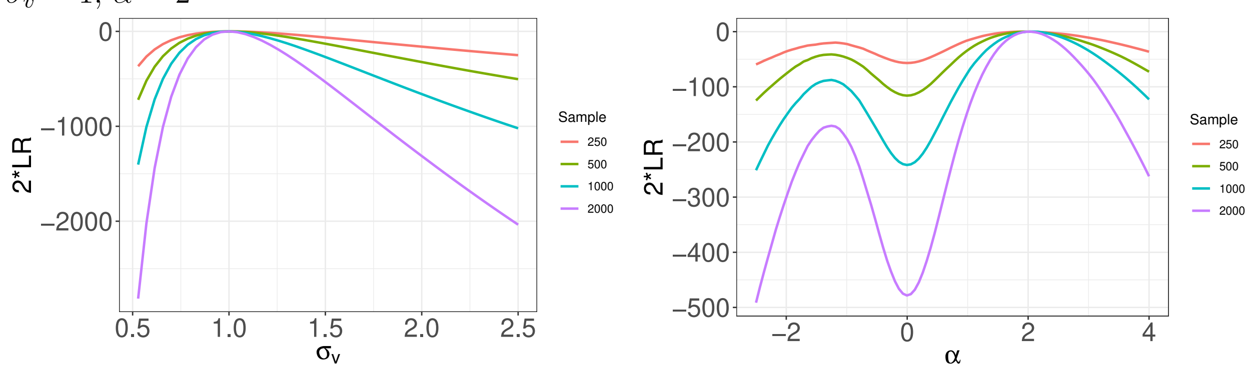


Figure C2: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

C.2 SN-Exp regression

Table C3: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v - u$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$, and $u \sim Exp(1/\sigma_u)$. Bias and MSE, $\sigma_v = 0.5$, $\sigma_u = 0.5$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_u		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 0.5$, $\alpha = 0.5$, $\sigma_u = 0.5$										
250	−0.0252	0.0083	0.0012	0.0052	−0.0214	0.0039	0.1066	0.0115	−0.9240	2.3132
500	−0.0155	0.0041	−0.0028	0.0026	−0.0129	0.0018	0.0876	0.0077	−0.5232	1.6280
1000	−0.0084	0.0019	0.0004	0.0012	−0.0047	0.0010	0.0626	0.0040	−0.2449	1.0230
2000	−0.0029	0.0009	−0.0005	0.0006	−0.0040	0.0005	0.0478	0.0025	−0.5033	0.6415
$\sigma_v = 0.5$, $\alpha = 1$, $\sigma_u = 0.5$										
250	−0.0257	0.0062	−0.0015	0.0044	−0.0190	0.0027	0.0506	0.0047	−0.3371	2.0782
500	−0.0086	0.0026	0.0007	0.0019	−0.0050	0.0013	0.0307	0.0024	−0.0945	1.0007
1000	−0.0024	0.0014	−0.0012	0.0010	−0.0025	0.0007	0.0123	0.0012	−0.0613	0.2820
2000	−0.0029	0.0007	−0.0002	0.0005	−0.0029	0.0003	0.0041	0.0009	−0.0530	0.1628
$\sigma_v = 0.5$, $\alpha = 2$, $\sigma_u = 0.5$										
250	−0.0238	0.0045	0.0041	0.0033	−0.0207	0.0020	−0.0053	0.0024	−0.4419	1.7232
500	−0.0101	0.0020	0.0009	0.0017	−0.0103	0.0009	−0.0052	0.0013	−0.1619	0.8080
1000	−0.0056	0.0011	−0.0001	0.0009	−0.0052	0.0004	−0.0047	0.0007	−0.1032	0.4184
2000	−0.0038	0.0006	−0.0002	0.0005	−0.0046	0.0003	−0.0037	0.0004	−0.0893	0.2173

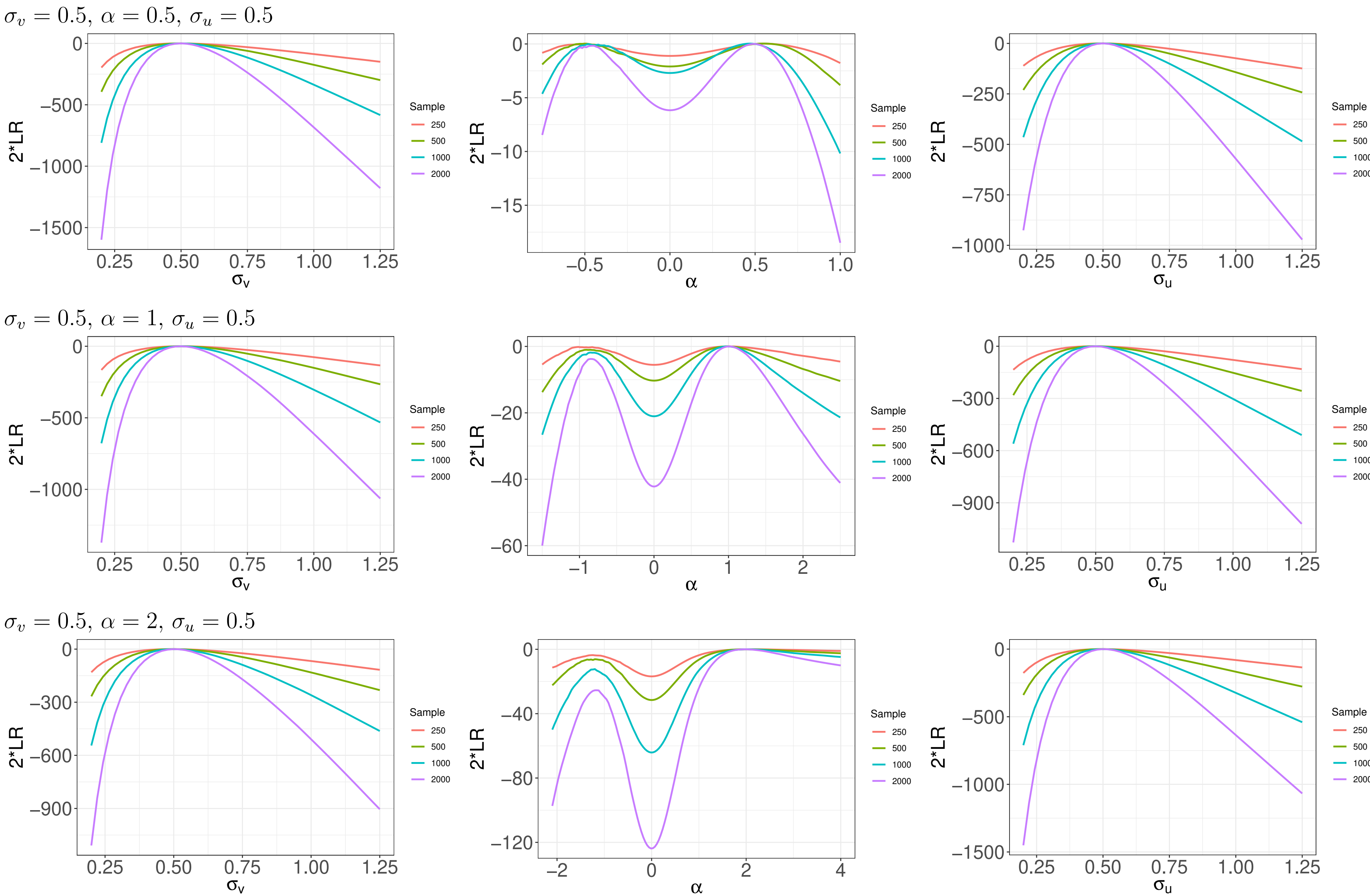


Figure C3: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

Table C4: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v - u$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$, and $u \sim Exp(1/\sigma_u)$. Bias and MSE, $\sigma_v = 0.5$, $\sigma_u = 1$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_u		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 0.5$, $\alpha = 0.5$, $\sigma_u = 1$										
250	−0.0275	0.0117	0.0034	0.0087	−0.0221	0.0056	0.1375	0.0189	−1.2614	3.6717
500	−0.0117	0.0049	0.0009	0.0043	−0.0116	0.0025	0.1081	0.0117	−0.9804	2.4087
1000	−0.0042	0.0023	0.0003	0.0020	−0.0057	0.0013	0.0811	0.0066	−0.7277	1.6318
2000	−0.0005	0.0012	−0.0018	0.0010	−0.0022	0.0007	0.0610	0.0038	−0.2050	0.9351
$\sigma_v = 0.5$, $\alpha = 1$, $\sigma_u = 1$										
250	−0.0272	0.0087	0.0021	0.0076	−0.0253	0.0049	0.0701	0.0079	−1.3022	4.8366
500	−0.0128	0.0040	0.0039	0.0037	−0.0110	0.0023	0.0472	0.0041	−0.3316	2.7463
1000	−0.0056	0.0019	0.0027	0.0016	−0.0039	0.0012	0.0293	0.0021	−0.0842	0.9196
2000	−0.0015	0.0010	−0.0007	0.0009	−0.0014	0.0006	0.0144	0.0011	−0.0029	0.2628
$\sigma_v = 0.5$, $\alpha = 2$, $\sigma_u = 1$										
250	−0.0240	0.0069	−0.0063	0.0060	−0.0274	0.0044	0.0064	0.0044	−0.9107	4.5340
500	−0.0158	0.0034	0.0034	0.0033	−0.0153	0.0018	−0.0055	0.0024	−0.4427	1.6533
1000	−0.0055	0.0015	0.0024	0.0015	−0.0052	0.0009	−0.0077	0.0011	−0.2135	0.8287
2000	−0.0046	0.0008	0.0005	0.0008	−0.0027	0.0004	−0.0038	0.0006	−0.0468	0.4180

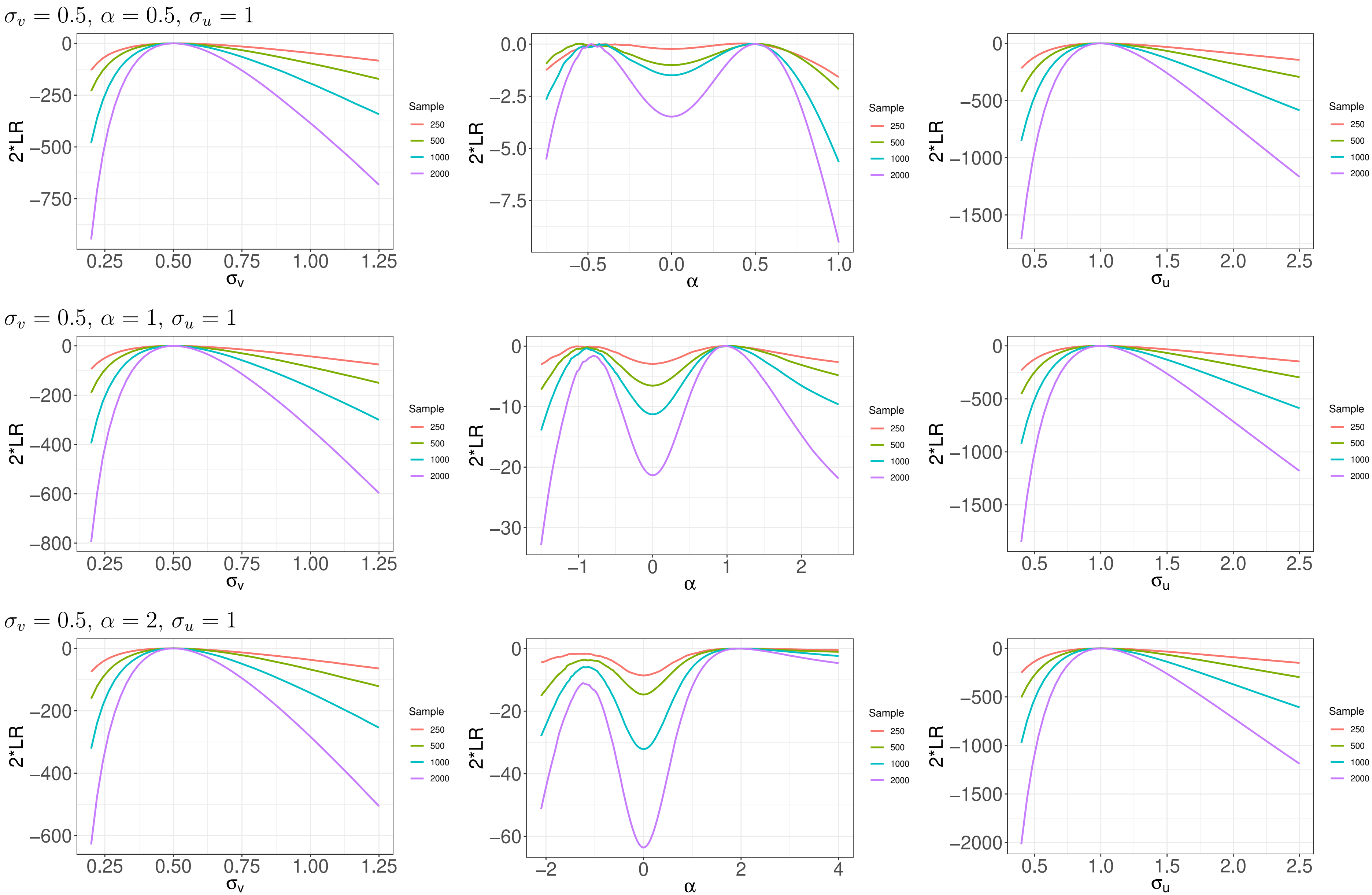


Figure C4: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

Table C5: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v - u$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$, and $u \sim Exp(1/\sigma_u)$. Bias and MSE, $\sigma_v = 1$, $\sigma_u = 0.5$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_u		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 1$, $\alpha = 0.5$, $\sigma_u = 0.5$										
250	−0.0211	0.0292	−0.0021	0.0152	0.0010	0.0185	0.1120	0.0162	−0.4927	0.9480
500	0.0004	0.0144	−0.0072	0.0065	0.0110	0.0084	0.0771	0.0090	−0.4190	0.4520
1000	−0.0102	0.0081	0.0004	0.0035	−0.0082	0.0047	0.0688	0.0063	−0.3222	0.2884
2000	−0.0111	0.0042	−0.0026	0.0019	−0.0078	0.0024	0.0539	0.0044	−0.0981	0.2498
$\sigma_v = 1$, $\alpha = 1$, $\sigma_u = 0.5$										
250	−0.0263	0.0200	0.0101	0.0128	0.0043	0.0115	0.0190	0.0100	−0.2292	0.9933
500	−0.0060	0.0107	0.0015	0.0060	0.0007	0.0050	0.0018	0.0056	−0.1091	0.4430
1000	−0.0091	0.0058	0.0001	0.0029	−0.0078	0.0030	0.0003	0.0038	−0.0616	0.1496
2000	−0.0053	0.0026	−0.0006	0.0014	−0.0065	0.0015	0.0026	0.0020	−0.0164	0.0731
$\sigma_v = 1$, $\alpha = 2$, $\sigma_u = 0.5$										
250	−0.0103	0.0122	−0.0039	0.0098	0.0090	0.0047	−0.0505	0.0078	−0.3012	1.0124
500	−0.0070	0.0070	0.0009	0.0046	0.0039	0.0021	−0.0243	0.0033	−0.1340	0.4523
1000	0.0033	0.0033	−0.0011	0.0025	0.0051	0.0013	−0.0123	0.0013	−0.0337	0.2051
2000	0.0002	0.0015	−0.0023	0.0012	0.0015	0.0006	−0.0050	0.0007	−0.0077	0.0955

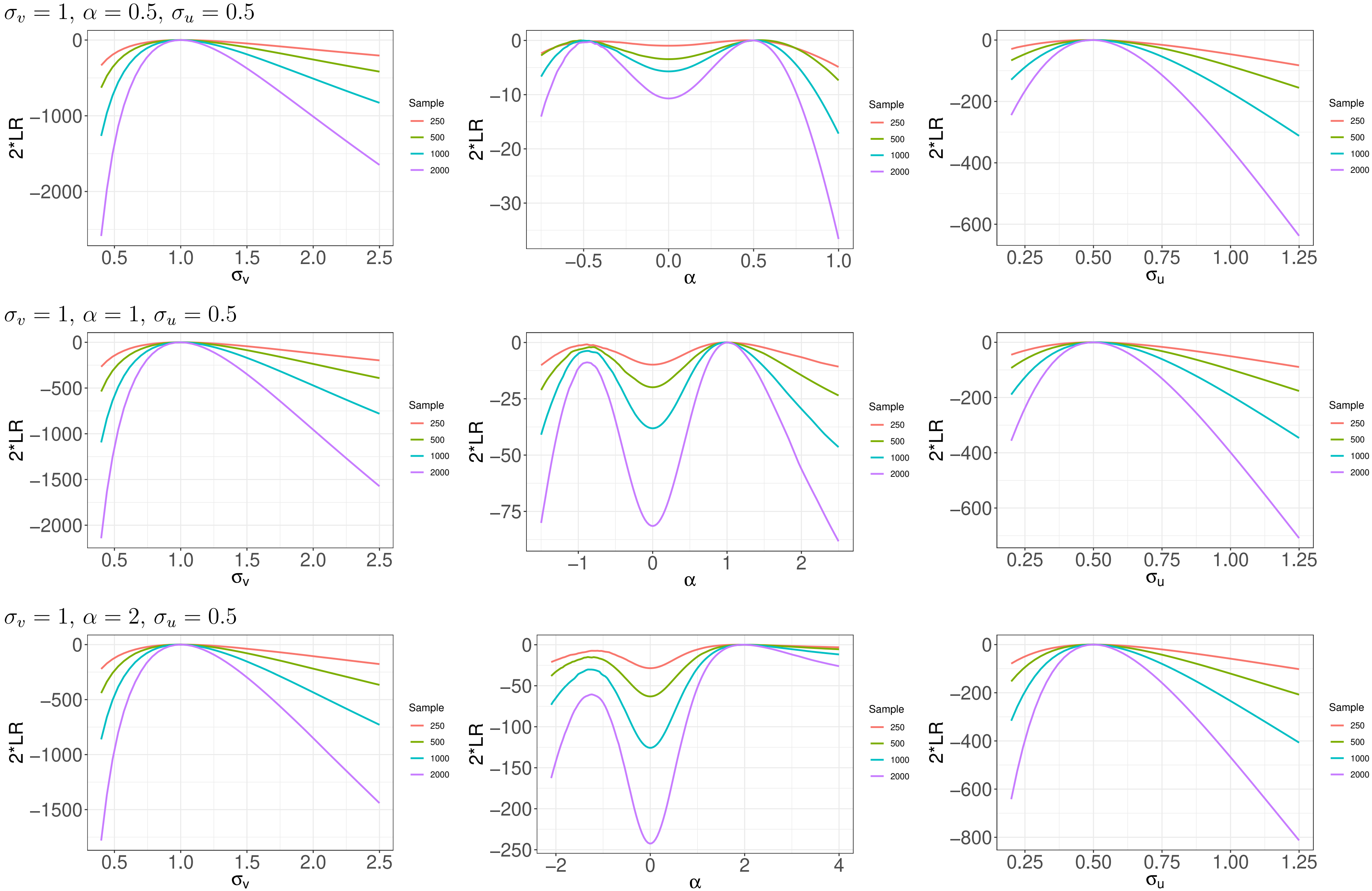


Figure C5: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

Table C6: The DGP is $\ln Y = \beta_0 + \beta_1 \ln X + v - u$, where $\beta_0 = 0.2$, $\beta_1 = 0.5$, $v \sim SN(0, \sigma_v^2, \alpha)$, and $u \sim Exp(1/\sigma_u)$. Bias and MSE, $\sigma_v = 1$, $\sigma_u = 1$, and three values of α, 0.5, 1, and 2

N	β_0		β_1		σ_u		σ_v		α	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_v = 1$, $\alpha = 0.5$, $\sigma_u = 1$										
250	−0.0639	0.0348	0.0034	0.0189	−0.0410	0.0157	0.2055	0.0430	−0.6357	2.2089
500	−0.0199	0.0152	0.0006	0.0102	−0.0216	0.0069	0.1611	0.0262	−0.4993	1.5535
1000	−0.0205	0.0072	0.0020	0.0049	−0.0136	0.0034	0.1260	0.0163	−0.5048	1.0149
2000	−0.0106	0.0040	−0.0012	0.0027	−0.0143	0.0018	0.0772	0.0081	−0.5175	0.5326
$\sigma_v = 1$, $\alpha = 1$, $\sigma_u = 1$										
250	−0.0479	0.0263	0.0128	0.0168	−0.0347	0.0109	0.0935	0.0175	−0.4009	2.5465
500	−0.0189	0.0116	−0.0010	0.0081	−0.0168	0.0052	0.0538	0.0091	−0.1998	1.0067
1000	−0.0170	0.0054	0.0046	0.0042	−0.0085	0.0027	0.0286	0.0055	−0.0846	0.3813
2000	−0.0094	0.0030	0.0000	0.0024	−0.0105	0.0016	−0.0168	0.0057	−0.2408	0.4031
$\sigma_v = 1$, $\alpha = 2$, $\sigma_u = 1$										
250	−0.0516	0.0187	0.0106	0.0161	−0.0388	0.0081	−0.0128	0.0104	−0.4755	1.6486
500	−0.0246	0.0086	0.0050	0.0071	−0.0218	0.0037	−0.0107	0.0051	−0.2244	0.9788
1000	−0.0164	0.0045	0.0014	0.0036	−0.0135	0.0019	−0.0159	0.0036	−0.1565	0.5443
2000	−0.0144	0.0023	0.0016	0.0018	−0.0116	0.0012	−0.0168	0.0020	−0.1553	0.3266

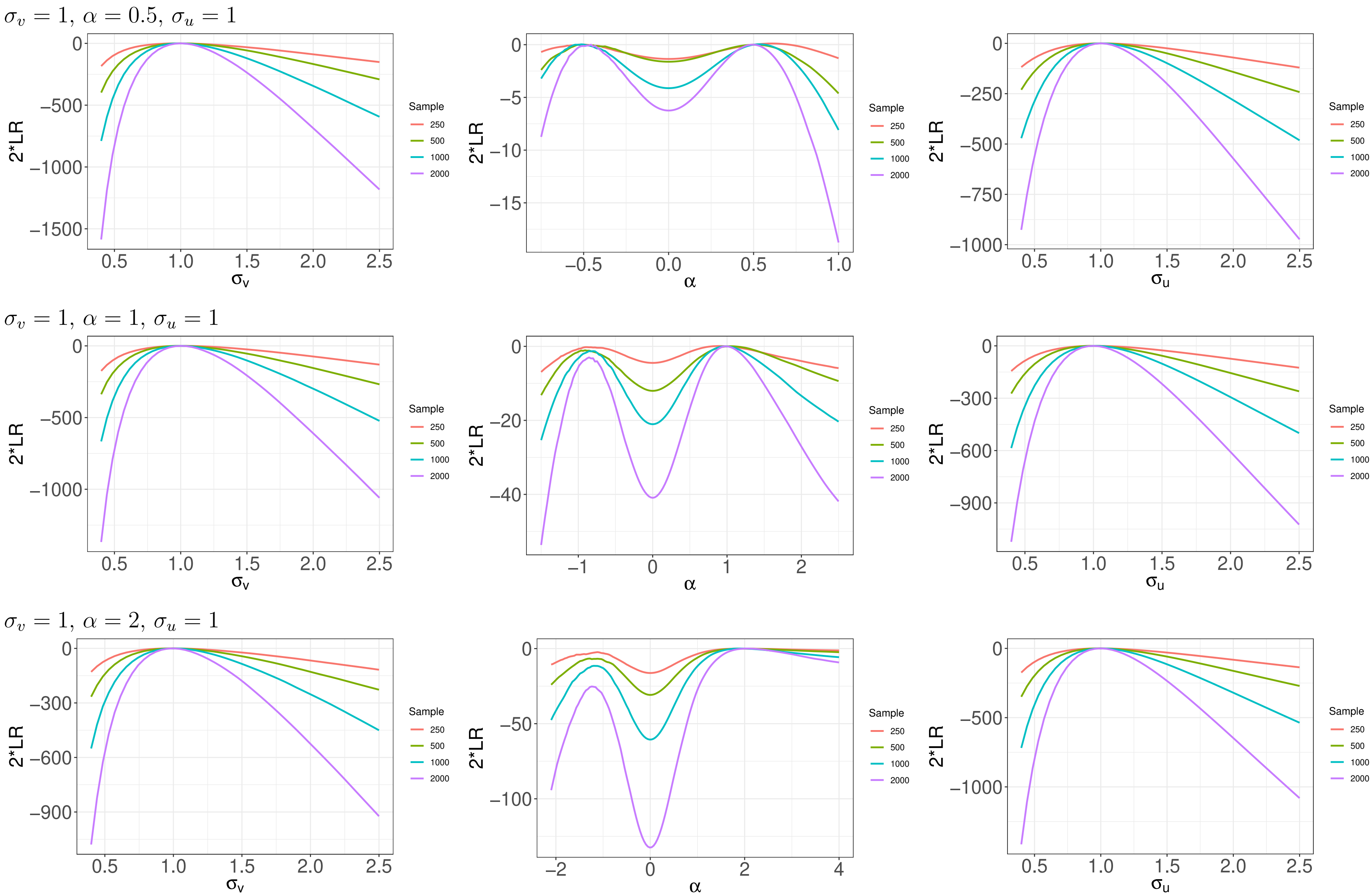


Figure C6: Median over 99 replications. Double difference between likelihood calculated at true values and likelihood calculated varying one parameter.

D Simulations and applications