ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 2/2022

Binary endogenous treatment in stochastic frontier models with an application to soil conservation in El Salvador

Samuele Centorrino, María Pérez Urdiales, Boris Bravo-Ureta, Alan Wall

Departamento de Economía

Universidad de Oviedo

Available online at: https://www.unioviedo.es/oeg/

BINARY ENDOGENOUS TREATMENT IN STOCHASTIC FRONTIER MODELS WITH AN APPLICATION TO SOIL CONSERVATION IN EL SALVADOR

SAMUELE CENTORRINO*

(*) *Corresponding Author.*
Economics Department, the State University of New York at Stony Brook, USA.
Email address: samuele.centorrino@stonybrook.edu.

MARÍA PÉREZ URDIALES

Economics Department, the State University of New York at Stony Brook, USA.
Email address: maria.perez-urdiales@stonybrook.edu.

BORIS BRAVO-URETA

Department of Agricultural and Resource Economics, University of Connecticut, USA.
Email address: boris.bravoureta@uconn.edu.

ALAN WALL

Oviedo Efficiency Group and Department of Economics, University of Oviedo, Spain.
Email address: awall@uniovi.es.

Abstract. Improving the productivity of the agricultural sector is part of one of the Sustainable Development Goals set by the United Nations. To this end, many international organizations have funded training and technology transfer programs that aim to promote productivity and income growth, fight poverty and enhance food security among small farmers in developing countries. Stochastic production frontier analysis can be a useful tool when evaluating the effectiveness of these programs. However, accounting for endogenous selection into treatment, often intrinsic to these interventions, only recently received any attention in the stochastic frontier literature. In this work, we extend the classical maximum likelihood estimation of stochastic production frontier models by allowing both the production frontier and inefficiency to depend on a potentially endogenous binary treatment. We use instrumental variables to define an assignment mechanism for the treatment, and we explicitly model the density of the first and second-stage composite error terms. We provide empirical evidence of the importance of controlling for endogeneity in this setting using farm-level data on a soil conservation program in El Salvador.

Keywords: Binary treatment; Endogeneity; Stochastic Frontier; Maximum Likelihood; Technical efficiency.

Date: June 19, 2022.
The authors wish to thank Christopher F. Parmeter, participants the 11th North American Productivity Workshop (NAPW XI), especially Subal C. Kumbhakar, and the Brown Bag Seminar at the Center of Productivity and Performance of Loughborough University for useful comments and remarks.

1. Introduction

The global need to increase productivity via technological change and efficiency improvements in the agricultural sector has been recognized in the United Nations 2030 Agenda for Sustainable Development. In particular, the Sustainable Development Goal (SDG) #2 aims to end hunger and improve the agricultural productivity and incomes of small-scale farmers while promoting resilient agricultural practices and sustainable food production systems. A growing number of governments, development organizations, and agencies are implementing programs targeting this goal. Many of these programs work at the scale of smallholder farm households and often include support for the adoption of innovative technologies and practices, as well as funding for technical assistance, agricultural education and training (see Bravo-Ureta et al., 2020, Jimi et al., 2019 and also de Janvry et al., 2017, for a review).

However, participation in these programs often occurs voluntarily, which may lead to endogenous selection of participants into treatment. Farmers who choose to participate (i.e., who self-select into the program) may share specific characteristics that distinguish them from non-participants. For instance, the participants' cultivated land may suffer more from erosion, and as a consequence, they may be less efficient than non-participants. If endogenous selection of participants into the treatment is not controlled for, one might conclude that the program is not effective because the agricultural efficiency of those who participate is lower than those who do not.

Stochastic Frontier Analysis is a popular method to assess agricultural efficiency as a measure of the production potential of the agricultural sector. The production frontier is defined as the quantity of output that can be produced with some input mix, given the technology and the environment. Efficiency (or inefficiency) is measured by the distance of each producer from the frontier, and it is usually modeled using a one-sided unobserved random variable. Similarly, the output may be measured with an error, or there may be other sources of variation in the outcome not observed by the econometrician that result in a further stochastic component. This *composite* error term is a defining feature of stochastic frontier models. In this framework, endogenous selection has been modeled either by considering the dependence between program participation and the two-sided stochastic component of the error term (Greene, 2010); or by considering the dependence between program participation and the unobserved inefficiency component of error term (Kumbhakar et al., 2009). However, as pointed out by Parmeter and Kumbhakar (2014, Section 6.3.3, p.293), “a framework where selection is based on [the composite error term] does not exist." We fill this gap in the current literature and develop a framework in which selection is based on the composite error term.

In our framework, we let the treatment be binary and denote it with a random variable Z. The decision of farmer i to participate in the treatment can be described as an index function, Z_i^*, which depends on observed ($\tilde{W}_i$) and unobserved (η_i) variables (Heckman and Robb, 1985). If the index Z_i^* is greater than a threshold (often normalized to 0), then the farmer decides to participate in the treatment and $Z_i = 1$; otherwise, she does not and $Z_i = 0$. The decision of farmers to participate in the treatment generates dependence between unobservable individual characteristics (i.e., preferences and/or managerial abilities) and the treatment assignment. If the unobservable factors, η_i, are correlated with the composite error term in the stochastic frontier framework, then the latter affects the choice of the farmer to overtake the treatment. Hence, the composite error term *confounds* the effect of Z_i on the output and on the mean of the inefficiency term. This endogeneity problem renders standard estimation methods inconsistent. As mentioned above, the challenge is to provide a framework which allows for the treatment assignment to be potentially correlated with both components of the composite error term.

In the stochastic frontier literature, the issue of endogeneity with respect to the two-sided error component has been previously treated as an issue of *sample selection bias*. The main rationale for this choice is that treated and non-treated units come from two distinct populations. The selection mechanism therefore applies as we only observe the treatment group in one sample and the control group in another sample (Greene, 2010; Lai, 2015). Bravo-Ureta et al. (2012) considers a framework similar to Greene (2010) to control for selection on unobservables, coupled with a propensity score matching technique to additionally control for selection on observables.

However, there are two reasons why the sample selection framework may not be appropriate to study endogenous selection into treatment. From a methodological perspective, the sample includes observations from both the treatment and the control group coming from the same population (e.g., the population of farmers in a specific region). In this context, there does not seem to be any particular reasons to estimate two different frontiers for a population of farmers based on whether they participate in the treatment or not. From a statistical perspective, the application of the sample selection approach requires splitting the full sample into two subsamples, which effectively reduces the degrees of freedom. Potential differences in output elasticities between treatment and control group can be appropriately introduced using interaction terms between the treatment variable and the inputs without splitting the sample, as we explain below.

An alternative approach is proposed by Kumbhakar et al. (2009). In particular, Kumbhakar et al.'s (2009) paper considers the endogeneity of technology choice (conventional or organic farming) by jointly estimating technology and technology choice using a single-step maximum likelihood method. In their framework, the technology choice directly depends on the inefficiency term, and therefore endogeneity operates through the inefficiency term only. Moreover, they do not consider treatment participation as a potential determinant of inefficiency.

The approaches in both Greene (2010) and Kumbhakar et al. (2009) are based on simulated maximum likelihood, while we provide a closed-form likelihood function, and therefore we do not need to resort to simulations. This has both computational advantages and may improve the finite sample properties of the estimator.

More recently, Chen et al. (2020) study a general model with binary endogenous treatment and mediator, that are potentially correlated with the composite error term. Their approach uses a propensity score assumption, which is used to construct moment conditions that are robust to the potential endogeneity of the treatment. However, they do not provide an estimator of technical efficiency.

To the best of our knowledge, there does not exist a maximum likelihood framework in stochastic frontier analysis which allows one to control for potential correlation between program participation and both the unobservable idiosyncratic component and the stochastic inefficiency.

Our contribution to this literature is to provide a model that allows one to control for endogeneity coming from both sources. We allow the treatment to enter the model in a flexible way, so that participation in the program can act as both a determinant of (ine)efficiency as well as a facilitating input. This is crucial as it permits to test whether program participation helps farmers produce more efficiently, given the technology; and/or modifies the technology. Our empirical strategy is to employ instrumental variables to construct an auxiliary assignment mechanism for program participation.

We then propose a maximum likelihood framework in which we jointly model the density of the first stage error and the density of the composite error term common to the stochastic production frontier. Under appropriate conditional independence assumptions, we derive the likelihood function in closed form, which allows us to use standard estimation and inference procedures, making the model straightforward to estimate and interpret. We also provide some theoretical results about identification and estimation, with a focus on the parameter capturing dependence between the stochastic inefficiency and the unobservable first-stage error. We show that only the magnitude of the dependence is identified, but not its sign. Moreover, when the true value of this parameter is 0, the information matrix is singular and the model is only second-order identified. As this limiting case is relevant for practitioners who wish to test the lack of endogeneity with respect to the inefficiency component, we discuss some testing procedures which can be applied in this context, and also discuss how to construct confidence intervals that are robust to the lack of first-order identification Rotnitzky et al. (2000); Andrews (2001); Bottai (2003); Ekvall and Bottai (2022).

Our framework is similar in spirit to Kumbhakar et al. (2009), in that we also use a single-step maximum likelihood method. However, we model the dependence of both components of the error term. Compared to the model proposed by Chen et al. (2020), we also impose explicit distributional assumptions on both the inefficiency term and the stochastic component. However, our approach is based on a one-step maximum likelihood estimator and allows one to obtain an estimator of technical efficiency for each producer, which is not provided in Chen et al. (2020). The ability to estimate technical efficiency is an essential feature of stochastic frontier models, as it allows comparisons across different observations (Farrell, 1957; Jondrow et al., 1982). We also contribute to the recent literature about endogeneity in stochastic frontier models (Amsler et al., 2016, 2017; Centorrino and Pérez-Urdiales, 2021), by studying the case in which the endogenous variable is binary.

We apply the proposed method to a sample of smallholder farm households from El Salvador. The data consist of a sample of participants in an environmental program promoting soil conservation practices, as well as a control group of non-participant farmers. In this empirical analysis, standard stochastic frontier estimation does not show any effect of the policy, either on the production level or on farmers' technical efficiency. By contrast, our approach reveals that program participation significantly improves technical efficiency. These results further highlight the need to control for endogeneity when evaluating such interventions, as this may substantially change the conclusions regarding their effectiveness.

The paper is structured as follows. In Section 2, we described the econometric model and our maximum likelihood estimator. Section 3 contain a finite sample assessment of our method, in which we also discuss the implementation of our estimator. Section 4 contains a description of the sample and outlines our empirical results. Finally, Section 5 concludes.

2. Binary Treatment in Stochastic Production Frontier

2.1. **Model.** We consider the following stochastic frontier regression model:

$$Y = m(X, Z, \beta) + V - U, \tag{1}$$

where Y is the logarithm of output; $m(X, Z, \beta)$ is the logarithm of the production frontier, which depends on some unknown parameter β, some production inputs, X, and other *environmental factors*, Z; and $\varepsilon = V - U$ is a composite error term. This error term is divided into two parts: V is a stochastic component with mean equal to 0, and $U \geq 0$ is an inefficiency term that captures the shortfall of the producer from the frontier. The latter may depend on other observed characteristics of producers (for instance, experience and education) that are often introduced as a scale factor affecting the distribution of U (Simar et al., 1994; Alvarez et al., 2006). Thus, we write $U = U_0 g(Z, \delta)$, where $g(\cdot, \cdot)$ is the so-called *scale function* which is specified by the econometrician and depends on some unknown parameter δ. This function is normalized such that $g(0, \delta) = 1$. For instance, in our empirical illustration, we take Z as the dummy for participation in the program fostering soil-conservation. This is a binary treatment variable that takes value 1 if the producer participates in the program and 0 otherwise. This variable can affect both the production frontier and the inefficiency of the producer. To simplify the discussion that follows, we assume that the participation dummy is the only *environmental factor*, so that Z is univariate. This specific model can be easily generalized when Z includes also other exogenous environmental factors.

As Z is binary, we can write the production frontier as

$$m(X, Z, \beta) = m(X, \beta_0) + Z m(X, \beta_1),$$

so that the frontier *shifts* from $m(X,\beta_0)$ to $m(X,\beta_0)+m(X,\beta_1)$ as the treatment variable changes from 0 to 1. Therefore, our model in (1) becomes

$$Y = m(X,\beta_0) + Zm(X,\beta_1) + V - U_0 g(Z,\delta).$$

For instance, in the prevalent case in which the logarithm of the production function is linear in parameters (e.g., Cobb-Douglas or translog), this modeling strategy involves the inclusion in the production frontier of the dummy variable for the treatment and the interaction between the treatment dummy and each one of the inputs (or a subset of these regressors). This specification is in line with McCloud and Kumbhakar (2008) in that the treatment may affect the output through the input coefficients (as a facilitating input that is not necessary for output production), the technological change parameter (generating a frontier shift), and the efficiency term.

Maximum likelihood estimation is a popular approach to obtain estimators of the parameters (β,δ) in a stochastic frontier framework (Kumbhakar and Lovell, 2003). Although heavily parametrized, the likelihood specification allows one to identify and estimate the variance of the inefficiency term. This in turn permits the construction of an estimator of technical (in)efficiency, which captures the distance of each farmer from the production frontier.

These maximum likelihood estimators can be based on a variety of assumptions about the distributions of V and U_0. However, the most popular model assumes that V follows a normal distribution and that U_0 follows a half-normal distribution (Aigner et al., 1977). Moreover, one usually assumes that V is independent of U_0 and that (X,Z) are fully independent of (V,U_0).

In our framework, the treatment is not taken to be independent of the joint error term (V,U_0). Volunteering for the treatment can depend both on the inefficiency of the producer and on other preferences that are unobserved to the econometrician. This implies that the treatment is endogenous. In this case, the stochastic frontier model based on an independence assumption between Z and (V,U_0) would lead to an inconsistent estimator of (β,δ).[1] Our goal is to construct a maximum likelihood estimator that generalizes the normal-half-normal stochastic frontier model when the treatment is allowed to be endogenous.

[1]Production inputs can also be correlated with the composite error term (Mundlak, 1961; Schmidt and Sickles, 1984). However, we focus here on the endogeneity of the treatment. Constructing an estimator that is also robust to endogeneity in the inputs is possible, although we do not tackle it in this paper (see Centorrino and Pérez-Urdiales, 2021).

In econometrics, the use of instrumental variables is a popular method to deal with endogeneity. That is, we assume there exists a vector of instruments, W, of dimension $q \geq 1$, which is correlated with Z but independent of (V, U_0). However, Z enters the second-stage equation nonlinearly, so the usual approach used in linear instrumental variable models of obtaining the predicted values of Z from the first stage and using them in a second stage maximum likelihood estimation instead of Z would not lead to a consistent estimation of (β, δ) (Wooldridge, 2015; Amsler et al., 2016). An alternative approach is based on a so-called *control function* assumption. That is, one can write $Z = \mathbb{1}\left(\tilde{W}\gamma + \eta \geq 0\right)$, where $\tilde{W} = (W, X)$, and assume that all the dependence between Z and (V, U_0) is captured by η. The latter can be considered an omitted variable in the second stage. Thus, once we control for η, the dependence between Z and (V, U_0) disappears (Newey et al., 1999; Imbens and Newey, 2009).

In particular, we use a Probit specification to model the treatment assignment (i.e., the first stage equation). Thus, we have that

$$P\left(Z = 1|\tilde{W}\right) = 1 - \Phi\left(-\tilde{W}\gamma\right),$$

where $\Phi(\cdot)$ is the cdf of a standard normal distribution. The main assumptions of the Probit model are that $\eta \sim N(0,1)$ and that $\tilde{W}$ is independent of η. A maximum likelihood framework requires specification of the dependence between η and (V, U_0). More formally, this is done by modeling the conditional density of (V, U_0) given η. Consistently with the work of Centorrino and Pérez-Urdiales (2021), we assume that any dependence between V and U_0 has to happen through η. When there is no endogeneity, this assumption is equivalent to the full independence between V and U_0 usually imposed in stochastic frontier models.

Our main assumptions are formalized as follows.

Assumption 2.1. *(i)* $\tilde{W} \perp\!\!\!\perp (V, \eta, U_0)$ *and* $V \perp\!\!\!\perp U_0|\eta$.

(ii) $Z = \mathbb{1}\left(\tilde{W}\gamma + \eta \geq 0\right)$, *with* $\eta \sim N(0,1)$.

(iii) $V|\eta \sim N\left(\rho_V \sigma_V \eta, \sigma_V^2(1 - \rho_V^2)\right)$.

(iv) $U_0|\eta \sim FN\left(\rho_U \sigma_U \eta, \sigma_U^2(1 - \rho_U^2)\right)$, *where FN denotes a folded normal distribution.*

The parameters ρ_V and ρ_U capture the dependence between (V, U_0) and η, respectively. The conditional pdf of U_0 given η is written as

$$f_{U_0|\eta}(u|\eta) = \frac{1}{\sqrt{2\pi(1-\rho_U^2)\sigma_U^2}}\left\{\exp\left(-\frac{(u-\rho_U\sigma_U\eta)^2}{2(1-\rho_U^2)\sigma_U^2}\right)+\exp\left(-\frac{(u+\rho_U\sigma_U\eta)^2}{2(1-\rho_U^2)\sigma_U^2}\right)\right\}. \quad (2)$$

Centorrino and Pérez-Urdiales (2021) have shown that this specification of the conditional density of U_0 provides a generalization with endogeneity to the normal half-normal stochastic frontier model (Aigner et al., 1977). It can be seen that the pdf in Equation (2) reduces to the half-normal distribution when $\rho_U = 0$, i.e., when the treatment is assigned independently of the efficiency of the producer. In the following, we refer to ρ_U as a dependence parameter.

Maximum likelihood estimation in stochastic frontier models is usually based on the density of the composite error term $\varepsilon = V - U$. In the case where the treatment is endogenous, our maximum likelihood estimator is based on the joint density of (ε, η), which can be decomposed into the product of the conditional density of ε given η and the marginal density of η which, in our case, is a standard normal density.

From Centorrino and Pérez-Urdiales (2021), the conditional density of ε given η is equal to

$$\begin{aligned}
f_{\varepsilon|\eta}(\varepsilon|\eta) &= \int f_{V|\eta}(\varepsilon+u|\eta)\,(g(Z,\delta))^{-1} f_{U_0|\eta}((g(Z,\delta))^{-1}u|\eta)du \\
&= \frac{1}{\sqrt{2\pi}\tilde{\sigma}(Z)}\left\{\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)}+\frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)}-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta+\rho_U\sigma_U(Z)\eta)^2}{2\tilde{\sigma}^2(Z)}\right)\right. \\
&\quad \left. +\ \Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)}-\frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)}-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta-\rho_U\sigma_U(Z)\eta)^2}{2\tilde{\sigma}^2(Z)}\right)\right\},
\end{aligned}$$

where

$$\begin{aligned}
\sigma_U^2(Z) &= \sigma_U^2\,(g\,(Z,\delta))^2\,, \qquad \tilde{\sigma}_U^2(Z) = (1-\rho_U^2)\sigma_U^2(Z), \qquad \tilde{\sigma}_V^2 = (1-\rho_V^2)\sigma_V^2, \\
\tilde{\sigma}^2(Z) &= \tilde{\sigma}_U^2(Z)+\tilde{\sigma}_V^2, \qquad \lambda(Z) = \frac{\tilde{\sigma}_U(Z)}{\tilde{\sigma}_V}.
\end{aligned}$$

Remark 1. The scaling property of U is essential to derive the conditional density of ε given η. Given that $U_0 \perp\!\!\!\perp Z|\eta$, the scaling properties allows Centorrino and Pérez-Urdiales (2021) to write the conditional density of U given η as a scaled transformation of the density of U_0 given η.

Finally, using our assumption that η follows a standard normal distribution, the joint density of (ε, η) can be written as

$$
\begin{aligned}
f_{\varepsilon,\eta}(\varepsilon,\eta) =& \frac{1}{2\pi\tilde{\sigma}(Z)}\left\{\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)} + \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)} - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta+\rho_U\sigma_U(Z)\eta)^2}{2\sigma^2} - \frac{\eta^2}{2}\right)\right.\\
&\left. + \Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)} - \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)} - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta-\rho_U\sigma_U(Z)\eta)^2}{2\tilde{\sigma}^2(Z)} - \frac{\eta^2}{2}\right)\right\}.
\end{aligned}
$$

Let $\theta = (\beta', \delta', \sigma_U^2, \sigma_V^2, \rho_V, \rho_U, \gamma')'$ be the vector of parameters of interest. When Z is continuous, at least for identification purposes, we can assume that η is observed and thus define the likelihood using the joint density of ε and η obtained above (Centorrino and Pérez-Urdiales, 2021). When Z is binary, this is not possible, as the first stage error term, η, cannot be estimated from the data. We thus need to define the joint likelihood differently.

In similar frameworks (e.g., Probit and Logit models), the observable random variable is discrete, and we usually express the likelihood (conditional on exogenous covariates) as the cdf of a latent error term which follows a known distribution. In our case, we have two observable endogenous variables (Y, Z), and the likelihood is obtained by their density, conditional on the exogenous components, $\tilde{W}$. We aim at rewriting this density in terms of the error components (ε, η). Therefore, as η is latent, the likelihood is written with respect to its cdf. In particular, we aim at writing the likelihood as the product between the cdf of η conditional on ε and the pdf of ε.

To this end, we first consider the following joint probability of the observable endogenous variables. For $Z = 0$, we have

$$
\begin{aligned}
P(Y \leq y, Z = 0|\tilde{W} = \tilde{w}) =& P(m(x,0,\beta) + \varepsilon \leq y, Z = 0|\tilde{W} = \tilde{w})\\
=& P(\varepsilon \leq y - m(x,0,\beta), \eta \leq -\tilde{w}\gamma) = F_{\varepsilon,\eta}\left(y - m(x,0,\beta), -\tilde{w}\gamma\right),
\end{aligned}
$$

where the second line follows from the assumption of independence between (ε, η) and $\tilde{W}$. A similar derivation holds when $Z = 1$.

If we take the derivative of the joint probability in equation (3) with respect to its first argument, we obtain a function which is a pdf with respect to ε and a cdf with respect to η. In particular, we have

$$
\partial_1 F_{\varepsilon,\eta}\left(y - m(x,0,\beta), -\tilde{w}\gamma\right) = \int_{-\infty}^{-\tilde{w}\gamma} f_{\varepsilon|\eta}(y - m(x,0,\beta), \eta)d\eta
$$

$$= \int_{-\infty}^{-\tilde{w}\gamma} f_{\varepsilon|\eta}(y - m(x, 0, \beta)|\eta)\phi(\eta)d\eta,$$

where the second line follows from the Assumption that $\eta \sim N(0,1)$, and ϕ is the pdf of a standard normal distribution.

The likelihood function can thus be obtained as

$$\mathcal{L}(\theta) = \left(\int_{-\tilde{W}\gamma}^{\infty} f_{\varepsilon|\eta}(Y - m(X, Z, \beta)|\eta)\phi(\eta)d\eta \right)^{Z} \left(\int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon|\eta}(Y - m(X, Z, \beta)|\eta)\phi(\eta)d\eta \right)^{1-Z}, \quad (3)$$

where θ is defined above.

The integrals appearing in the likelihood function can be solved analytically. In particular, we obtain that the conditional cdf of η given ε is a mixture of two conditional skew-normal distributions (Azzalini and Dalla Valle, 1996; Azzalini, 2013). For $j = \{1, 2\}$, we let

$$\begin{aligned}
\Psi_{0,j}(Z,\theta) =& \Phi_2\left(\frac{-\tilde{W}\gamma - \mu_{\eta,j}(Z)\left(Y - m(X,Z,\beta)\right)}{\sigma_{\eta,j}(Z)}, \tau_j(Z)\left(Y - m(X,Z,\beta)\right); \rho^*(Z) \right), \\
\Psi_{1,j}(Z,\theta) =& \Phi\left(\tau_j(Z)\left(Y - m(X,Z,\beta)\right)\right) \\
& - \Phi_2\left(\frac{-\tilde{W}\gamma - \mu_{\eta,j}(Z)\left(Y - m(X,Z,\beta)\right)}{\sigma_{\eta,j}(Z)}, \tau_j(Z)\left(Y - m(X,Z,\beta)\right); \rho^*(Z) \right),
\end{aligned}$$

where $\Phi_2(\cdot,\cdot;\rho^*)$ is the cdf of a bivariate normal random variable with correlation parameter $\rho^*(Z)$, and $\{\mu_{\eta,j}(Z), \sigma_{\eta,j}(Z), \tau_j(Z), \rho^*(Z), \sigma^2_{\varepsilon,j}(Z)\}$ are functions of the parameter θ, whose dependence is suppressed for simplicity. We are finally able to show that the likelihood function can be written as

$$\begin{aligned}
\mathcal{L}(\theta) = & \left(\sum_{j=1,2} \Psi_{1,j}(Z,\theta) \frac{1}{\sigma_{\varepsilon,j}(Z)} \phi\left(\frac{Y - m(X,Z,\beta)}{\sigma_{\varepsilon,j}(Z)} \right) \right)^{Z} \times \\
& \left(\sum_{j=1,2} \Psi_{0,j}(Z,\theta) \frac{1}{\sigma_{\varepsilon,j}(Z)} \phi\left(\frac{Y - m(X,Z,\beta)}{\sigma_{\varepsilon,j}(Z)} \right) \right)^{1-Z}. \quad (4)
\end{aligned}$$

A detailed derivation is provided in Appendix A.1.

When $\rho_V = \rho_U = 0$,

$$\begin{aligned}
\sigma^2_{\varepsilon,1}(Z) =& \sigma^2_{\varepsilon,2}(Z) = \sigma^2_V + \sigma^2_U(Z) \\
\Psi_{1,1}(Z,\theta) =& \Psi_{1,2}(Z,\theta) = 1 - \Phi(-\tilde{W}\gamma)
\end{aligned}$$

$$\Psi_{0,1}(Z,\theta) = \Psi_{0,2}(Z,\theta) = \Phi(-\tilde{W}\gamma),$$

and the likelihood reduces to the product of the pdf of a skew-normal distribution (the pdf of ε) and the cdf of a normal distribution (the cdf of η), which would be the likelihood function if the composite error term is independent of Z. This would be the standard approach in Stochastic Frontier Analysis (Kumbhakar and Lovell, 2003). Also, if $\rho_V = 0$, and we assume that $\sigma_U^2(Z)$ is constant wrt Z, we can write $\eta_i = e_i - \rho_U^2 U_{0i}/(\sigma_V^2 + \sigma_U^2)$, and our model will collapse to the one proposed by Kumbhakar et al. (2009).

2.2. **Identification.** Let $\ell(\theta) = \log \mathcal{L}(\theta)$ be the log-likelihood function, and assume that $E[|\ell(\theta)|] < \infty$ for all $\theta \in \Theta$. As we can restrict Θ to be a compact parameter space, and the likelihood function is continuous in θ, there exists at least one solution to the maximization of the log-likelihood function (Gourieroux and Monfort, 1995).

We focus our identification analysis on the parameter ρ_U. To this end, we maintain the following assumption.

Assumption 2.2. *Let* $\theta_1 = (\beta', \delta', \sigma_U^2, \sigma_V^2, \rho_V, \gamma')'$. *The matrix*

$$E\left[\nabla^2_{\theta_1\theta_1'}\ell(\theta_0)\right]$$

is negative definite and has full rank.

This assumption imposes that the parameter θ_1 is first-order locally identified (Sargan, 1983). In particular, we require that the variance of the inefficiency term $\sigma_{U,0}^2 > 0$. Lee and Chesher (1986) and Lee (1993) have shown that when $\sigma_{U,0}^2 = 0$, the stochastic frontier model is not first-order identified. Moreover, in our model, whenever $\sigma_U^2 = 0$, (δ, ρ_U) are not identified. We believe this case is worthy of future investigation, but we rule it out here for simplicity.

Proposition 2.1. *Let Assumptions 2.1-2.2 hold, and* $\rho_{U,0}$ *to be such that*

$$E\left[\frac{\partial \ell(\theta_{1,0}, \rho_{U,0})}{\partial \rho_U}\right] = 0.$$

We have that

(i) $E\left[\frac{\partial \ell(\theta_{1,0}, -\rho_{U,0})}{\partial \rho_U}\right] = 0.$

(ii) $\frac{\partial \ell(\theta_1,0)}{\partial \rho_U} = 0$, *for any* θ_1, *and*

$$E\left[\frac{\partial^2 \ell(\theta_{1,0},0)}{\partial \rho_U^2}\right] = 0.$$

This Proposition extends the result shown in Centorrino and Pérez-Urdiales (2021) to the case in which the endogenous variable is binary. Part (i) states that if $\rho_{U,0}$ is a solution of the maximization problem, so is $-\rho_{U,0}$. That is, the sign of ρ_U is not identified. Part (ii) states that $\rho_U = 0$ always satisfies the first order conditions of the maximization problem for any value of θ_1. This result implies that the matrix of second derivatives has rank equal to $dim(\theta) - 1$, and the model is not first-order identified at $\rho_U = 0$. A proof of this Proposition is provided in Appendix A.2.[2]

This identification issue is illustrated in Centorrino and Pérez-Urdiales (2021), and it follows because the likelihood function is an even function of ρ_U and symmetric about $\rho_U = 0$. In our setting, dealing with this identification issue is easier, compared to the framework in Centorrino and Pérez-Urdiales (2021), as the parameter ρ_U is scalar.

We thus restrict the support of ρ_U to be $[0,1]$ (Sundberg, 1974*a*). We let $\bar{\Theta}$ to be the parameter's space which embeds the restriction on ρ_U, and we define

$$\theta_0 = \arg\max_{\theta \in \bar{\Theta}} E\left[\ell(\theta)\right], \tag{5}$$

which exists and is (locally) unique under Assumption 2.2.

2.3. **Estimation and Inference.** For estimation, we consider an iid sample from the joint distribution of (Y, X, Z, W), which we denote $\{(Y_i, X_i, Z_i, W_i), i = 1, \dots, n\}$, where each observation obeys to the model in (1).

Let $\ell_n(\theta) = \log(\mathcal{L}_n(\theta))$, with

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n}\left(\sum_{j=1,2}\Psi_{1,j}(\theta,Z_i)\frac{1}{\sigma_{\varepsilon,j}(Z_i)}\phi\left(\frac{Y_i - m(X_i,Z_i,\beta)}{\sigma_{\varepsilon,j}(Z_i)}\right)\right)^{Z_i} \times \left(\sum_{j=1,2}\Psi_{0,j}(\theta,Z_i)\frac{1}{\sigma_{\varepsilon,j}(Z_i)}\phi\left(\frac{Y_i - m(X_i,Z_i,\beta)}{\sigma_{\varepsilon,j}(Z_i)}\right)\right)^{1-Z_i}. \tag{6}$$

[2]A similar identification problem arises in Zero Inefficiency Stochastic Frontier models, see Kumbhakar et al. (2013); Rho and Schmidt (2015).

Estimation is straightforward, with the maximum likelihood estimator of the parameter θ given by

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \ell_n(\theta),$$

We analyze our estimator's asymptotic distribution depending on the true value of the parameter ρ_U. To simplify our analysis, we make the following high-level assumption.

Assumption 2.3. $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Under Assumptions 2.2 and 2.3 and since the log-likelihood function is at least twice continuously differentiable with respect to the parameter θ_0, when ρ_U is strictly in the interior of $[0,1]$, standard theory of maximum likelihood estimation applies and we can claim that

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \mathcal{I}_{\theta_0}^{-1}\right),$$

where $\mathcal{I}_{\theta_0}$ is the Fisher's information matrix.

However, the asymptotic distribution and the rate of convergence of our estimator are non-standard when $\rho_{U,0}$ is equal to 0. In this case, it follows from the result of Proposition 2.1 that we have a singular Hessian matrix, and one of the parameters of interest is at the boundary of the parameter space. This implies that we do not have the standard $\sqrt{n}$-rate of convergence, and that our estimator of ρ_U is not asymptotically normal (Sundberg, 1974*b*; Andrews, 1999; Rotnitzky et al., 2000). However, a reparametrization of the log-likelihood function allows us to obtain the rate of convergence and asymptotic distribution of our estimator.

Let $\rho_{2,U} = \rho_U^2$. The following theorem gives the asymptotic properties of our estimator when $\rho_{U,0} = 0$.

Theorem 2.1. *Let Assumptions 2.1-2.2 hold with $\rho_{U,0} = 0$, and $(Z_{\theta_1}, Z_{\rho_{2,U}})$ a normal random vector such that $dim(Z_{\theta_1}) = dim(\theta_1)$, $dim(Z_{\rho_{2,U}}) = 1$, with covariance matrix $\mathcal{I}_1^{-1}$, where $\mathcal{I}_1^{-1}$ is the inverse of*

$$\mathcal{I}_1 = \begin{bmatrix} \mathcal{I}_{\theta_1} & \mathcal{I}'_{\theta_1 \rho_{2,U}} \\ \mathcal{I}_{\theta_1 \rho_{2,U}} & \mathcal{I}_{\rho_{2,U}} \end{bmatrix}.$$

Further let $\hat{\tau}_{\rho_{2,U}} = \max\{Z_{\rho_{2,U}}, 0\}$. Then

(i)

$$\sqrt{n}\begin{pmatrix}\hat{\theta}_1 - \theta_{1,0} \\ \hat{\rho}_{2,U}\end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_{\theta_1} - \mathcal{I}_{\theta_1}^{-1}\mathcal{I}'_{\theta_1\rho_{2,U}}\hat{\tau}_{\rho_{2,U}} \\ \hat{\tau}_{\rho_{2,U}}\end{pmatrix}$$

(ii) $n^{1/4}\hat{\rho}_U = O_P(1)$.

This theorem reflects existing results in statistics about estimation of a parameter that is not first-order identified (Rotnitzky et al., 2000). As the order of identification in our model is even, the marginal asymptotic distribution of $\hat{\rho}_{2,U}$ at $\rho_{U,0} = 0$ is an equal mixture of a point-mass at 0, and a half-normal distribution (Chernoff, 1954; Andrews, 1999). The result in Part (ii) is a direct consequence of Part (i). However, it is worth highlighting that our estimator has rates of convergence slower than $\sqrt{n}$, and may not be asymptotically normal, depending on the true value of the parameter ρ_U. These results have important implications for conducting inference and constructing confidence intervals for ρ_U.

In particular, an important hypothesis to be tested is whether Z is independent of the inefficiency term, i.e., $\rho_U = 0$. The trinity of tests is an obvious candidate, but the implementation of these tests is not straightforward because of the non-standard asymptotic properties of the estimator of ρ_U.

Andrews (2001) studies the properties of the trinity of tests when some parameters are at the boundary. His theoretical results about the Likelihood Ratio (LR) test can be used following Theorem 2.1. The critical values from the asymptotic distribution of the LR statistic are obtained from an equal mixture between a mass point at 0 and a χ^2 distribution with one degree of freedom (see Self and Liang, 1987; Andrews, 1999; Rotnitzky et al., 2000; Andrews, 2001).

Obtaining uniformly valid standard errors and confidence intervals is more cumbersome. Our results suggest that one can first construct a test of the null hypothesis that $\rho_U = 0$. If the null is rejected, we are then able to use standard errors and confidence intervals based on the normal approximation. If the null cannot be rejected, then other methods should be used to obtain confidence intervals.

Bottai (2003) has shown that, when there is a singularity in the Fisher Information matrix, confidence intervals based on inverting the Likelihood Ratio or Wald test fail to have nominal coverage near the point of singularity. However, an appropriately modified version of the score test can be used to obtain confidence intervals with asymptotically uniform nominal coverage. This theoretical result holds, however, only when θ is a scalar parameter. More recently, Ekvall and Bottai (2022) have provided a generalization for any multivariate parameter θ and when the rank deficiency of the information matrix is potentially larger than one. This modified version of the score test in a model, such as ours, that is second-order identified, is constructed using the second derivative of the likelihood function at the point of singularity, normalized by the expected value of its square. This methodology could be potentially extended to the model we study in this paper to construct uniformly valid confidence intervals when $\rho_{U,0} = 0$. A formal theoretical result is deferred to future research.

2.4. **Technical Efficiency.** To complete our framework, we obtain a feasible estimator of technical efficiency, $TE_i = \exp(-U_i)$. Researchers are often interested in obtaining the technical efficiency for each producer. Amsler et al. (2017) and Centorrino and Pérez-Urdiales (2021) obtain an estimator of this quantity from the conditional distribution of U given ε and η. However, as η is not observed in our case, we have to follow the standard approach and obtain the estimator of technical inefficiency from the conditional distribution of U given ε. The latter distribution can be derived as

$$f_{U|\varepsilon}(u|\varepsilon) = \int f_{U|\varepsilon,\eta}(u|\varepsilon,\eta) f_{\eta|\varepsilon}(\eta|\varepsilon) d\eta.$$

Details about the exact computations of this density are given in Appendix A.1. We let

$$\begin{aligned}
\sigma_{1\star} &= \frac{\tilde{\sigma}_V \tilde{\sigma}_U(Z)}{\sigma(Z)} \sqrt{1 + \frac{q_1^2(Z)\tilde{\sigma}^2(Z)}{\tilde{\sigma}^2(Z) + \rho_1^2(Z)}} \\
\sigma_{2\star} &= \frac{\tilde{\sigma}_V \tilde{\sigma}_U(Z)}{\sigma(Z)} \sqrt{1 + \frac{q_2^2(Z)\tilde{\sigma}^2(Z)}{\tilde{\sigma}^2(Z) + \rho_2^2(Z)}} \\
\mu_{1\star} &= -\frac{\tilde{\sigma}_V \tilde{\sigma}_U(Z)}{\sigma(Z)} \varepsilon \left(\frac{\lambda(Z)}{\sigma(Z)} - \frac{q_1(Z)\rho_1(Z)}{\tilde{\sigma}^2(Z) + \rho_1^2(Z)} \right) \\
\mu_{2\star} &= -\frac{\tilde{\sigma}_V \tilde{\sigma}_U(Z)}{\sigma(Z)} \varepsilon \left(\frac{\lambda(Z)}{\sigma(Z)} - \frac{q_2(Z)\rho_2(Z)}{\tilde{\sigma}^2(Z) + \rho_2^2(Z)} \right),
\end{aligned}$$

where the definition of the other parameters is given in the Appendix, and the dependence on the variable Z has been removed for simplicity.

We obtain that

$$E\left[\exp(-U)|\varepsilon\right] = \omega_1(\varepsilon)\exp\left(-\mu_{1\star} + \frac{\sigma_{1\star}^2}{2}\right)\frac{1-\Phi\left(-\frac{\mu_{1\star}}{\sigma_{1\star}} + \sigma_{1\star}\right)}{\Phi\left(\tau_1(Z)\varepsilon\right)}$$
$$+ \omega_2(\varepsilon)\exp\left(-\mu_{2\star} + \frac{\sigma_{2\star}^2}{2}\right)\frac{1-\Phi\left(-\frac{\mu_{2\star}}{\sigma_{2\star}} + \sigma_{2\star}\right)}{\Phi\left(\tau_2(Z)\varepsilon\right)}, \quad (7)$$

where $\omega_l(\varepsilon)$, with $l = 1, 2$ are weights, such that $\omega_1(\varepsilon) + \omega_2(\varepsilon) = 1$, whose closed-form expression is given in the Appendix. Finally, the mean technical efficiency (Lee and Tyler, 1978) can be obtained as

$$E\left[\exp(-U)\right] = E\left[E\left[\exp(-U)|\varepsilon\right]\right],$$

by the law of iterated expectations.

3. Simulations

We replicate a similar simulation scheme as in Amsler et al. (2017) and Centorrino and Pérez-Urdiales (2021). We consider the following model

$$Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_1 + V_i - U_{0i}\exp\left(Z_{1i}\delta_1 + Z_{2i}\delta_2\right),$$

with $\beta_0 = 0$ and $\beta_1 = \beta_2 = 0.66074$, $\delta_1 = 0$ and $\delta_2 = 0$ and where the random variables (X_{1i}, X_{2i}, Z_{1i}) are exogenous (i.e. fully independent of the composite error term), and Z_{2i} is our endogenous treatment variable. We consider two instruments (W_{1i}, W_{2i}), also fully independent of the error term, and such that

$$Z_{2i} = \mathbb{1}\left(\gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 Z_{1i} + \gamma_4 W_{1i} + \gamma_5 W_{2i} + \eta \geq 0\right),$$

where $\mathbb{1}(\cdot)$ is the indicator function, $\gamma_0 = -0.1$, $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_5 = 0.31623$, and $\gamma_6 = 1$.

The exogenous variables (X_1, X_2, Z_1, W_1, R) are generated from a joint normal distribution with mean equal to 0 and covariance matrix with diagonal elements equal to 1, and off-diagonal elements equal to 0.5. $W_2 = \mathbb{1}(R > 0.5)$.

We generate the pair (V,η) from the normal distribution

$$\begin{pmatrix} V_i \\ \eta_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right),$$

so that $\rho_V = 0.5$.

The stochastic inefficiency term is generated as

$$U_0 = \sigma_U |\rho_U \eta + \sqrt{1-\rho_U^2}\epsilon|,$$

where $\sigma_U^2 = \pi/(\pi-2)$, and $\epsilon \sim N(0,1)$.

We consider two simulation schemes that differ according to the value of the parameter ρ_U. In Setting 1, U_0 is independent of η (i.e. $\rho_U = 0$). In Setting 2, $\rho_U = 0.5$. Sample sizes are fixed to $n = \{250, 500, 1000\}$, and we run 1000 replications for each scenario.

The implementation of our estimator is rather straightforward. The starting point is the log-likelihood function from equation (6)

$$\begin{aligned} \ell_n(\theta) = \sum_{i=1}^{n} \Bigg[& Z_{2i} \log \left(\sum_{j=1,2} \Psi_{1,j}(Z_i,\theta) \frac{1}{\sigma_{\varepsilon,j}(Z_i)} \phi\left(\frac{Y_i - m(X_i, Z_i, \beta)}{\sigma_{\varepsilon,j}(Z_i)} \right) \right) \\ & + (1 - Z_{2i}) \log \left(\sum_{j=1,2} \Psi_{0,j}(Z_i,\theta) \frac{1}{\sigma_{\varepsilon,j}(Z_i)} \phi\left(\frac{Y_i - m(X_i, Z_i, \beta)}{\sigma_{\varepsilon,j}(Z_i)} \right) \right) \Bigg], \end{aligned}$$

which is maximized numerically with respect to the parameter θ.

An essential step of numerical optimization procedures is to select a starting value. As far as the parameters (β,δ) are concerned, an initial value can be chosen by estimating a stochastic frontier model which does not account for endogeneity. Similarly, an initial value for the parameter γ can be obtained by a Probit regression of Z_2 on the instruments and all the other exogenous variables included in the model. Regarding the parameters (ρ_V,ρ_U), their starting values are taken by uniform draws from the interval $[-1,1]\times[0,1]$. As convergence to a local maximum might be an issue, we draw several random points around the starting value and initialize the search at every one of these points. While computationally expensive, this procedure is more robust to local maxima.

In parametric models with endogeneity, it is also common practice to perform estimation in two steps: first one obtains an estimator of the parameter γ from a Probit model, and then one estimates the remaining parameters by holding $\hat{\gamma}$ fixed. While this procedure is computationally more efficient, as it reduces the dimension of the parameters' space, we recommend against it. First of all, there is no guarantee that this procedure yields a numerically equivalent estimator to the maximization of the log-likelihood with respect to the full parameters' space. Moreover, standard errors of the two-step procedure are invalid, and one still needs to obtain the numerical Hessian matrix of the full likelihood to obtain valid standard errors. Our implementation is thus based on the maximization of the log-likelihood with respect to the full parameter vector, θ.

We report results of these simulations in Tables 1 and 2 below. For each parameter, we report the mean and standard deviation of the estimator computed over the simulated samples. As the parameter ρ_U is constrained to be positive, we report its median rather than its mean in Setting 1. According to our results, the sampling distribution of $\hat{\rho}_U$ should have a mass at zero with probability 0.5, when $\rho_{U,0} = 0$. Hence, the median should get closer to 0 as the sample size increases.

		$n = 250$		$n = 500$		$n = 1000$	
	TRUE	Location	Std. Dev.	Location	Std. Dev.	Location	Std. Dev.
β_0	0.0000	-0.1280	0.3557	-0.0553	0.2103	-0.0202	0.1259
β_1	0.6607	0.6744	0.1126	0.6684	0.0768	0.6635	0.0524
β_2	0.6607	0.6734	0.1138	0.6622	0.0768	0.6617	0.0516
δ_1	0.0000	-0.0128	0.3456	-0.0065	0.0703	-0.0015	0.0414
δ_2	0.0000	0.2783	2.7451	-0.0230	1.7010	0.0282	0.4486
σ_U^2	2.7519	2.3015	1.2517	2.5231	0.8242	2.6500	0.5426
σ_V^2	1.0000	1.1318	0.4058	1.0517	0.2439	1.0210	0.1514
$\rho_{U,\eta}$	0.0000	0.0019	0.2936	0.0018	0.2143	0.0031	0.1606
$\rho_{V,\eta}$	0.5000	0.5426	0.1810	0.5220	0.1255	0.5112	0.0810
γ_0	-0.1000	-0.1033	0.1100	-0.1007	0.0782	-0.1015	0.0545
γ_1	0.3162	0.3224	0.1390	0.3223	0.0929	0.3175	0.0670
γ_2	0.3162	0.3354	0.1405	0.3203	0.0921	0.3214	0.0653
γ_3	0.3162	0.3245	0.1359	0.3212	0.0913	0.3224	0.0651
γ_4	0.3162	0.3305	0.1306	0.3220	0.0910	0.3191	0.0612
γ_5	1.0000	1.0293	0.2482	1.0141	0.1786	1.0119	0.1219

TABLE 1. *Location and Standard Deviation for Setting 1*

Our estimator behaves as expected. We also notice a smaller finite-sample bias in Setting 2. This is due to the slower rate of convergence of $\hat{\rho}_U$ is Setting 1, which may affect the estimation of the other parameters of the model.

Finally, we report summary statistics for our estimators of technical efficiency using the Battese-Coelli formula provided in Equation 7. To give a reference point to the reader, in both simulation

	TRUE	n = 250 Mean	Std. Dev.	n = 500 Mean	Std. Dev.	n = 1000 Mean	Std. Dev.
β_0	0.0000	-0.1216	0.3377	-0.0518	0.2027	-0.0153	0.1154
β_1	0.6607	0.6647	0.1108	0.6659	0.0761	0.6627	0.0526
β_2	0.6607	0.6670	0.1112	0.6603	0.0756	0.6603	0.0522
δ_1	0.0000	-0.0084	0.2548	-0.0046	0.1025	0.0002	0.0379
δ_2	0.0000	-0.0821	2.9661	0.0193	0.2742	0.0062	0.1238
σ_U^2	2.7519	2.4028	1.3044	2.5898	0.9034	2.7088	0.6034
σ_V^2	1.0000	1.1299	0.4208	1.0464	0.2533	1.0117	0.1583
$\rho_{U,\eta}$	0.5000	0.4753	0.3072	0.4670	0.2218	0.4871	0.1364
$\rho_{V,\eta}$	0.5000	0.5106	0.2194	0.5074	0.1456	0.5005	0.0932
γ_0	-0.1000	-0.1029	0.1110	-0.1003	0.0787	-0.1011	0.0543
γ_1	0.3162	0.3224	0.1381	0.3234	0.0929	0.3175	0.0668
γ_2	0.3162	0.3342	0.1395	0.3204	0.0917	0.3213	0.0648
γ_3	0.3162	0.3244	0.1349	0.3214	0.0901	0.3216	0.0645
γ_4	0.3162	0.3324	0.1296	0.3212	0.0908	0.3205	0.0605
γ_5	1.0000	1.0391	0.2496	1.0178	0.1768	1.0150	0.1219

TABLE 2. *Mean and Standard Deviation for Setting 2*

schemes the mean technical efficiency is equal to

$$E\left[\exp(-U)\right] = 0.3847.$$

	N = 250 $\rho_U = 0$	$\rho_U = 0.5$	N = 500 $\rho_U = 0$	$\rho_U = 0.5$	N = 1000 $\rho_U = 0$	$\rho_U = 0.5$
Min.	0.000	0.000	0.001	0.001	0.001	0.000
1st Qu.	0.285	0.285	0.267	0.266	0.257	0.255
Median	0.449	0.450	0.417	0.419	0.404	0.405
Mean	0.439	0.435	0.403	0.402	0.391	0.390
3rd Qu.	0.578	0.577	0.542	0.542	0.530	0.530
Max.	1.000	1.000	1.000	1.000	1.000	0.834

TABLE 3. *Summary measures for the estimator of technical efficiency*

4. Soil Conservation in El Salvador

4.1. **Data and Model Specification.** We consider data from the *Programa Ambiental de El Salvador* or PAES, an environmental program promoting crop diversification and soil conservation practices. The dataset consists of a sample of PAES participants and a control group of nonparticipating farmers. The dataset consists of a sample of PAES participants and a control group of non-participating farmers.

The target population of this program was farmers with incomes below the poverty line and producing mostly staple crops, such as corn and beans. The program consisted in promoting soil conservation technologies among participants. The initial fieldwork took place in 2002, and a random sample of participants was re-surveyed in 2005, along with a sample of farmers who never received PAES benefits. Figure 1 shows the cantons (administrative divisions in El Salvador) where participants and non-participants are located (in black). For more details on the program and data collection, see Bravo-Ureta et al. (2006).

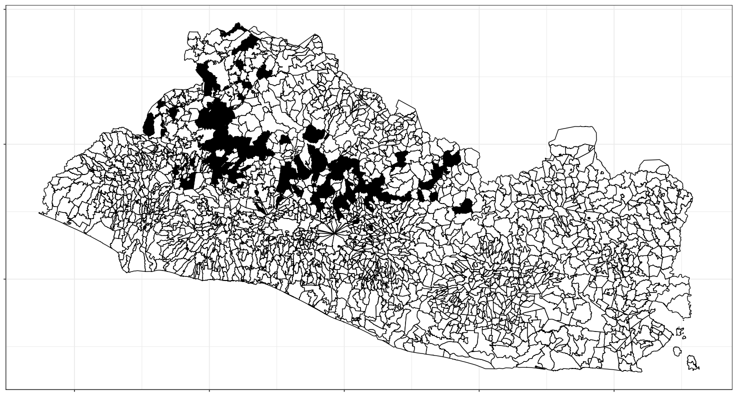

FIGURE 1. El Salvador - Location of cantons

For this application, we consider the cross-section of farmers surveyed in 2005 and specify the following model:

$$Y_i = \beta_0 + X_i\beta_1 + Z_{2i}\beta_2 + X_iZ_{2i}\beta_{1,2} + V_i - U_{0i}\exp(Z_{1i}\delta_1 + Z_{2i}\delta_2 + Z_{1i}Z_{2i}\delta_{1,2}), \qquad (8)$$

for $i = 1, \ldots, n$, with $\beta = (\beta_1', \beta_2', \beta_{1,2}')'$, $\delta = (\delta_1', \delta_2', \delta_{1,2}')'$, and where Y_i is the total value of *production* (measured in dollars); X_i is a vector of inputs including, *Labor* (number of hired and household laborers), *Land* (total area cultivated in manzanas where one manzana=0.7 has), *Fertilizers* (measured in dollars), *Pesticides* (measured in dollars) and *Seeds* (measured in dollars). Output and inputs are expressed in logs. $Z_i = (Z_{1i}, Z_{2i})$ is a vector of environmental factors, which is decomposed into two sub-vectors: Z_{2i} can act both as a frontier shifter as well as an inefficiency determinant, while Z_{1i} only contains inefficiency determinants. We consider *Participation* (dummy variable taking value 1 if the farmer is participating in PAES, 0 otherwise) as the frontier shifter, Z_{2i}. Z_{1i}, instead, includes the log of *Land*; a *Tenure* dummy equal to 1 if the farmer owns the entirety of the cultivated land, and 0 otherwise; the interaction between the latter two variables; a *Age* dummy equal to 1 if the farmer is 60 and older and 0 otherwise; *Education* (measured in years); a *No income* dummy, equal to 1 if the farmer has no outside sources of income and 0 otherwise; a *Foot*

Access dummy, equal to 1 that indicates whether farmers have only walking access to their plot; a *Car Access* dummy equal to 1 that indicates whether farmers also have access by car to their plot; and a risk diversification index, *Risk div*. This is a continuous index which compares the relative diversification of each farmer with respect to the average farmer. A higher value of the *Risk div* index implies higher risk diversification. Our indicator ranges over the entire real line.[3] We allow for first-order interaction terms between Z_1 and Z_2 to account for potential observed treatment heterogeneity and also for potential observable differences between the control and the treatment group, as alluded to in the Introduction. The reason to include the size of cultivated land into the inefficiency determinant is that, given soil erosion, which is common in El Salvador because of the frequent natural disasters, a larger plot requires the implementation of several soil conservation measures which are more costly and time-consuming to implement. The ability of the farmer to effectively implement these measures may also depend on the ownership structure. For this reason, we interact the total size of the plot with *Tenure*.

Finally, V_i is the idiosyncratic error and U_{0i} is the stochastic inefficiency term. Our goal is to estimate the parameters $\{\beta, \delta, \sigma_V^2, \sigma_U^2, \rho_U, \rho_V, \gamma\}$. The total sample size is equal to $n = 459$.

As stated earlier, in this application, *Participation* is an endogenous binary treatment variable. El Salvador was hit by a major earthquake in 2001 (the year prior to the beginning of the PAES program). We use this as an exogenous shock to determine treatment assignment. We construct a binary instrument, *Dist Earthquake*, which is equal to one if the log-distance (measured in kilometers) from the epicenter of the earthquake in 2001 was higher than 3.5, and 0 otherwise.[4] Our vector of instruments, W_i, also includes: *Electricity*, the proportion of families in the farmer's canton with access to electricity; *Bathroom*, the proportion of families in the farmer's canton with access to private bathrooms; *Wage Canton*, the average hourly wage in the canton where the farmer is located; and we also use three dummies for unobserved region characteristics that could have affected the

[3]We use the Ogive index of risk diversification in Wasylenko and Erickson (1978). The index is constructed as follows

$$\textit{Risk Div} = -\log\left(\sum_{j=1}^{C_i} \frac{(s_j - \bar{s}_j)^2}{\bar{s}_j}\right),$$

where s_j is the proportion of land devoted by the farmer to crop j, $\bar{s}_j$ is the average sample proportion of land devoted to crop j, and C_i is the total number of crops cultivated by farmer i.

[4]The density of *Dist Earthquake* is bimodal, with a mode around 1.5 and another one around 3.5. Since the percentage of farmers for whom *Dist Earthquake* is lower than 2.5 is slightly above 10%, we choose 3.5 as our cutoff point.

willingness of the farmer to participate in the program. In the following, we let $\tilde{W}_i = (X_i, Z_{1i}, W_i)$, to be the vector of included and excluded endogenous variables.

We test the specification of the assignment equation (the regression of the treatment variable on all the other exogenous regressors) as a Probit model using the test in Wilde (2008). We are not able to reject the null hypothesis that the error term follows a normal distribution at any standard significance level, which indicates that our assignment equation is not misspecified (value of the test statistic is 0.042, with a p-value equal to 0.979). Moreover, to assess the relevance of the instruments, we construct a likelihood ratio test which compares the unrestricted Probit model, with a Probit model where all coefficients associated with the instruments are set to zero. The null hypothesis of the test is that the instruments do not jointly have a significant effect on participation. The value of the likelihood ratio statistic is 128.76 which leads to a rejection of the null hypothesis with a p-value equal to 0. The estimated coefficients for the first-stage equation are given in the Appendix B.

When we ignore selection into treatment, we obtain the MLE assuming that $(U_0, V) \perp\!\!\!\perp (X, Z)$ with $V \sim N(0, \sigma_V^2)$ and $U_0 \sim N^+(0, \sigma_U^2)$. We remind the reader that the estimator which controls for potential treatment endogeneity would reduce to the former setting when $\rho_V = \rho_U = 0$.

4.2. **Results.** Table 4 reports results for our empirical example. We report our point estimates along the 95% confidence intervals. Confidence intervals for model that assumes exogeneity are constructed using a profile-likelihood method (Cox and Hinkley, 1979). This is because the sample size is relatively small and the normal approximation implicit in the more standard Wald-type confidence intervals may fail to hold. For the estimator that controls for potential endogeneity, we first perform a test for the absence of dependence between stochastic inefficiency and participation. We find that the null hypothesis of no dependence, i.e., $H_0 : \rho_U = 0$, cannot be rejected either using the LR test á la Andrews (2001) or the modified score test proposed by Bottai (2003) and Ekvall and Bottai (2022). At the 10% level, the critical value of the LR test is equal to 1.853, and our test statistic is numerically indistinguishable from 0, while the score test statistic is equal to 0 for a critical value of 2.706. This implies that we can exclude any dependence between farmers' stochastic inefficiency and their participation in the program. Hence, we only report results for the endogenous model when ρ_U is restricted to be equal to 0. When ρ_U is left unrestricted, point estimators of all

	Exogeneity			Endogeneity, $\rho_U = 0$		
	Estimate	CI		Estimate	CI	
β_0	3.7411	[3.0764	4.4057]	3.7137	[3.0045	4.2909]
β_{Land}	0.2249	[0.1157	0.3379]	0.2438	[0.1444	0.3457]
β_{Labor}	0.1934	[0.1191	0.2672]	0.1959	[0.1328	0.2568]
$\beta_{Fertilizer}$	0.1020	[0.0263	0.1772]	0.1045	[0.0412	0.1889]
$\beta_{Pesticides}$	0.0507	[-0.0111	0.1124]	0.0492	[0.0046	0.1038]
β_{Seeds}	0.1332	[0.0764	0.1898]	0.1362	[0.0872	0.2021]
$\beta_{Land \times Z_2}$	0.1197	[-0.0501	0.3557]	0.2127	[0.0670	0.3584]
$\beta_{Labor \times Z_2}$	-0.1370	[-0.2433	-0.0304]	-0.1512	[-0.2396	-0.0571]
$\beta_{Fertilizer \times Z_2}$	-0.1126	[-0.2377	0.0131]	-0.1054	[-0.2217	0.0017]
$\beta_{Pesticides \times Z_2}$	0.1266	[0.0289	0.2242]	0.1389	[0.0570	0.2182]
$\beta_{Seeds \times Z_2}$	0.0881	[-0.0126	0.1886]	0.1119	[0.0306	0.1906]
β_{Z_2}	0.6653	[-0.3423	1.6773]	0.2109	[-0.7217	1.1657]
δ_{Land}	0.0237	[-0.3282	0.4237]	0.0289	[-0.3754	0.3015]
δ_{Tenure}	-0.4883	[-1.5927	0.0953]	-0.6085	[-1.4837	0.2488]
$\delta_{Land \times Tenure}$	-0.1403	[-0.6168	0.4787]	-0.0917	[-0.3940	0.5131]
δ_{Age}	0.4653	[-0.0001	1.2973]	0.4745	[-0.0023	1.0874]
δ_{Educ}	-0.1750	[-0.4942	0.0573]	-0.2197	[-0.5570	0.0074]
$\delta_{Noincome}$	0.1079	[-0.3846	0.8403]	0.1810	[-0.3027	0.8214]
$\delta_{Footaccess}$	-1.0207	[-2.3106	-0.2441]	-1.1837	[-2.1660	-0.6023]
$\delta_{Caraccess}$	-0.4876	[-1.2807	0.2038]	-0.5814	[-1.2383	-0.0847]
$\delta_{Riskdiv}$	-0.2126	[-0.5129	0.0160]	-0.2203	[-0.4150	-0.0189]
$\delta_{Land \times Z_2}$	-0.2187	[-0.9204	1.2145]	0.2572	[-0.1642	0.6617]
$\delta_{Tenure \times Z_2}$	0.6209	[-0.3149	1.9683]	0.5870	[-0.3339	1.7936]
$\delta_{Land \times Tenure \times Z_2}$	0.3449	[-0.3869	1.0337]	0.3639	[-0.2355	0.8222]
$\delta_{Age \times Z_2}$	-0.6115	[-1.5179	-0.0256]	-0.6974	[-1.4997	-0.1376]
$\delta_{Educ \times Z_2}$	0.2659	[-0.0097	0.6346]	0.2583	[-0.0101	0.6826]
$\delta_{Noincome \times Z_2}$	-0.1531	[-0.9411	0.4498]	-0.1997	[-0.9787	0.4204]
$\delta_{Footaccess \times Z_2}$	0.7343	[-0.3806	2.1142]	1.0512	[0.4079	1.8286]
$\delta_{Caraccess \times Z_2}$	0.4337	[-0.5857	1.4887]	0.9611	[0.4319	1.5592]
$\delta_{Riskdiv \times Z_2}$	0.1211	[-0.1667	0.4449]	0.2663	[0.0265	0.5145]
δ_{Z_2}	-0.7030	[-5.8932	0.8935]	-1.9518	[-3.3784	-0.8608]
$\rho_{U,\eta}$				0.0000	[0.0000	0.0138]
$\rho_{V,\eta}$				0.3254	[0.1474	0.4922]
σ^2_U	0.6925	[0.0564	6.4017]	0.7241	[0.0997	3.5055]
σ^2_V	0.0978	[0.0777	0.1223]	0.1133	[0.0845	0.1338]

TABLE 4. *Estimation of the efficiency frontier with and without accounting for endogeneity.*

the other parameters are numerically indistinguishable. Since we assume that $\rho_{U,0} = 0$, the 95% confidence intervals in the endogenous setting are constructed by inverting the score test proposed in Ekvall and Bottai (2022).

The first set of columns in Table 4 shows the estimation results assuming exogeneity(parameter estimates and confidence intervals). We note that there is no significant shift in the production frontier for participants. Also, for participants, the output elasticity of *Labor* is lower and that of *Pesticides* is higher. *Participation* has a negative effect on inefficiency (i.e., improves efficiency), although the estimator is not significantly different from zero at the 5% level, according to our profile likelihood confidence intervals. Regarding the other efficiency determinants, we find substantially the same results for both participants and non-participants, though we do note that participation seems to lead to a larger inefficiency reduction for older farmers.

The results controlling for endogeneity are reported in the last three columns of Table 4 (parameter estimates and confidence intervals). We observe that the qualitative results are broadly consistent with the model assuming exogeneity. Thus, participation does not lead to any significant shift in the production frontier, and for participants the output elasticity of *Labor* is lower and that of *Pesticides* is higher. In contrast, in the model controlling for endogeneity the output elasticity for *Seeds* is higher for participants. Also, in the model assuming exogeneity, *Participation* did not have a significant effect on inefficiency, whereas in the model controlling for endogeneity the negative effect on inefficiency is now significant at the 5% level. Some other differences between the models with regard to inefficiency are also worth highlighting. For example, *Car access* had no effect on efficiency in the model assuming exogeneity but reduces inefficiency on the endogeneity model. Similarly, risk diversification (*Riskdiv*) had no effect on efficiency in the model assuming exogeneity but improves efficiency in the endogeneity model. Interestingly, the positive effect of risk diversification on efficiency is found to be lower for participants. Diversification therefore seems to be a good strategy for improving efficiency, though this is more effective or important for non-participants.

Moreover, the endogenous model also detects a positive correlation between the idiosyncratic component of the error term and *Participation* since $\hat{\rho}_V = 0.325$. This points out to the fact that self-selection into this particular program may be related to a preference component rather than to inefficiency considerations.

Finally, we report in Figure 2 a kernel density estimator of inefficiencies for non-participant farmers (left panel) and participant farmers (right panel) using the models that do and do not control for endogeneity. The dotted gray line is the density of inefficiencies for the model assuming exogeneity, while the solid black line is the density of inefficiencies in the model controlling for endogeneity. The Figure clearly illustrates that the model controlling for endogeneity detects larger efficiency scores in the group of participants and slightly lower ones among non-participants. We test for the difference between the distribution in the efficiency scores for participants and non-participants using the first-order stochastic dominance test in Linton et al. (2005), which allows for serial dependence in the observations. In particular, for participants, we wish to test whether the distribution of efficiency scores predicted by the exogenous model first-order stochastically dominates the distribution of efficiency scores predicted by the endogenous model. That is, for a given value of technical efficiency, the probability of being below that value is higher for the exogenous model than for the endogenous one, which would imply that efficiency scores predicted by the endogenous model are consistently higher for participants. Similarly, for non-participants, we wish to test that the distribution of scores predicted by the endogenous model first-order stochastically dominates the distribution of scores predicted by the exogenous model. The test in Linton et al. (2005) is a Kolmogorov-Smirnov type-test. Under the null, there is no first-order stochastic dominance. Critical values for the test are obtained by subsampling (Politis and Romano, 1994; Linton et al., 2005). The value of the test statistic for participants is 2.796, and the value of the test statistic for non-participants is 1.857. We choose a grid of several subsampling sizes, and, we obtain a median critical value of 2.054 for participants and 1.792 for non-participants at the 5% level. The null of no first-order stochastic dominance is therefore rejected in both cases. We can conclude that not controlling for endogeneity in our application would lead to incorrect efficiency scores, with these being overestimated for non-participants and underestimated for participants.

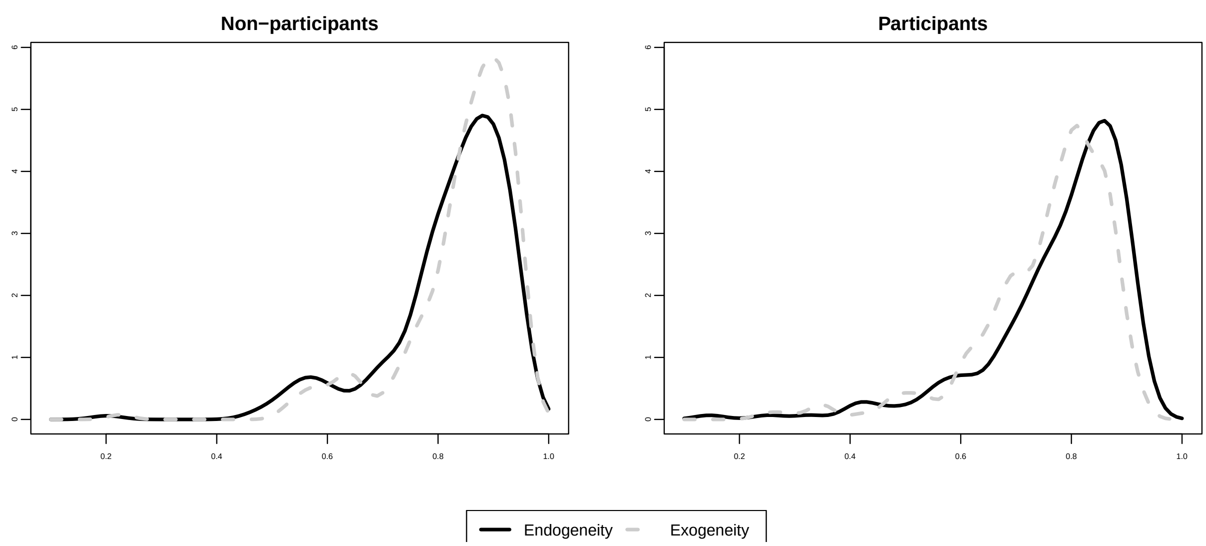

FIGURE 2. Density of estimated technical efficiency.

5. Conclusions

In the framework of policies aimed at increasing agricultural productivity, Stochastic Production Frontier Models can shed light on the effectiveness of programs targeting this goal. However, controlling for potential endogeneity associated with voluntary program participation is crucial to obtain an accurate estimate of the impact of the project.

In this paper, we propose a method to control for a binary endogenous treatment in Stochastic Production Frontier Models. In particular, we provide a simple maximum likelihood estimator based on distributional assumptions about the first and second-stage error terms. This estimator is in line with a more traditional approach to Stochastic Frontier Estimation, where one is usually interested in estimating the technical efficiency for each producer.

In the empirical application, we estimate Stochastic Production Frontiers for a sample of farmers in El Salvador participating in a soil conservation program and a control group of non-participant farmers. Our results show that the model assuming exogeneity does not detect any significant association between participation in the program and inefficiency. However, when we implement our methodology, we find that participation in the soil conservation program leads to an improvement in the level of technical efficiency.

The main policy implication of our study is that policymakers wishing to perform program evaluation in a Stochastic Frontier Model context should adequately control for endogeneity issues arising from voluntary program participation. Failure to adequately account for endogeneity may generating misleading conclusions about the effectiveness of such programs to generate improvements in productive efficiency. This is especially important in guiding future evidence-based policymaking in which the best possible use should be made of scarce resources.

An extension of our maximum likelihood approach to panel data models is a potential avenue for future research (Lai and Kumbhakar, 2018; Kutlu et al., 2019). In such a setting, the timing and evolution of the treatment over time, along with the assumptions one imposes on the composite error term, are paramount to understanding how our parametric approach applies, and whether it can allow for richer dependence between treatment and unobservables.

References

Aigner, D., Lovell, C. and Schmidt, P. (1977), 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics* **6**(1), 21 – 37.

Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006), 'Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics', *Journal of Productivity Analysis* **25**(3), 201–212.

Amsler, C., Prokhorov, A. and Schmidt, P. (2016), 'Endogeneity in stochastic frontier models', *Journal of Econometrics* **190**(2), 280 – 288.

Amsler, C., Prokhorov, A. and Schmidt, P. (2017), 'Endogenous environmental variables in stochastic frontier models', *Journal of Econometrics* **199**(2), 131 – 140.

Andrews, D. W. K. (1999), 'Estimation when a parameter is on a boundary', *Econometrica* **67**(6), 1341–1383.

Andrews, D. W. K. (2001), 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica* **69**(3), 683–734.

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

Azzalini, A. and Dalla Valle, A. (1996), 'The Multivariate Skew-Normal Distribution', *Biometrika* **83**(4), 715–726.

Bottai, M. (2003), 'Confidence regions when the Fisher information is zero', *Biometrika* **90**(1), 73–84.

Bravo-Ureta, B., Cocchi, H. and Solis, D. (2006), Adoption of Soil Conservation Technologies in El Salvador: A Cross-Section and Over-Time Analysis, Technical Report 18, Office of Evaluation and Oversight.

Bravo-Ureta, B. E., González-Flores, M., Greene, W. and Solís, D. (2020), 'Technology and Technical Efficiency Change: Evidence from a Difference in Differences Selectivity Corrected Stochastic Production Frontier Model', *American Journal of Agricultural Economics* **Forthcoming**.

Bravo-Ureta, B. E., Greene, W. and Solís, D. (2012), 'Technical efficiency analysis correcting for biases from observed and unobserved variables: an application to a natural resource management project', *Empirical Economics* **43**(1), 55–72.

Centorrino, S. and Pérez-Urdiales, M. (2021), 'Maximum Likelihood Estimation of Stochastic Frontier Models with Endogeneity', *Journal of Econometrics* **Forthcoming**.

Chen, Y.-T., Hsu, Y.-C. and Wang, H.-J. (2020), 'A Stochastic Frontier Model with Endogenous Treatment Status and Mediator', *Journal of Business & Economic Statistics* **38**(2), 243–256.

Chernoff, H. (1954), 'On the distribution of the likelihood ratio', *Ann. Math. Statist.* **25**(3), 573–578.

Cox, D. and Hinkley, D. (1979), *Theoretical Statistics*, Taylor & Francis.

de Janvry, A., Sadoulet, E. and Suri, T. (2017), Field Experiments in Developing Country Agriculture, *in* A. V. Banerjee and E. Duflo, eds, 'Handbook of Economic Field Experiments', Vol. 2 of *Handbook of Economic Field Experiments*, North-Holland, chapter 5, pp. 427 – 466.

Ekvall, K. O. and Bottai, M. (2022), 'Confidence Regions Near Singular Information and Boundary Points With Applications to Mixed Models', *Annals of Statistics* **Forthcoming**.

Farrell, M. J. (1957), 'The Measurement of Productive Efficiency', *Journal of the Royal Statistical Society. Series A (General)* **120**(3), 253–290.

Gourieroux, C. and Monfort, A. (1995), *Statistics and Econometric Models*, number v. 1 *in* 'Statistics and Econometric Models 2 volume set', Cambridge University Press.

Greene, W. (2010), 'A stochastic frontier model with correction for sample selection', *Journal of Productivity Analysis* **34**(1), 15–24.

Heckman, J. J. and Robb, R. (1985), 'Alternative methods for evaluating the impact of interventions: An overview', *Journal of Econometrics* **30**(1), 239 – 267.

Imbens, G. W. and Newey, W. K. (2009), 'Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity', *Econometrica* **77**(5), 1481–1512.

Jimi, N. A., Nikolov, P. V., Malek, M. A. and Kumbhakar, S. (2019), 'The effects of access to credit on productivity: separating technological changes from changes in technical efficiency', *Journal of Productivity Analysis* **52**(1), 37–55.

Jondrow, J., Lovell, C. K., Materov, I. S. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**(2), 233 – 238.

Kumbhakar, S. C., Parmeter, C. F. and Tsionas, E. G. (2013), 'A zero inefficiency stochastic frontier model', *Journal of Econometrics* **172**(1), 66 – 76.

Kumbhakar, S. C., Tsionas, E. G. and Sipiläinen, T. (2009), 'Joint estimation of technology choice and technical efficiency: an application to organic and conventional dairy farming', *Journal of Productivity Analysis* **31**(3), 151–161.

Kumbhakar, S. and Lovell, C. (2003), *Stochastic Frontier Analysis*, Stochastic Frontier Analysis, Cambridge University Press.

Kutlu, L., Tran, K. C. and Tsionas, M. G. (2019), 'A time-varying true individual effects model with endogenous regressors', *Journal of Econometrics* **211**(2), 539 – 559.

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

Lai, H.-p. (2015), 'Maximum likelihood estimation of the stochastic frontier model with endogenous switching or sample selection', *Journal of Productivity Analysis* **43**(1), 105–117.

Lee, L.-F. (1993), 'Asymptotic Distribution of the Maximum Likelihood Estimator for a Stochastic Frontier Function Model with a Singular Information Matrix', *Econometric Theory* **9**(3), 413–430.

Lee, L.-F. and Chesher, A. (1986), 'Specification testing when score test statistics are identically zero', *Journal of Econometrics* **31**(2), 121 – 149.

Lee, L.-F. and Tyler, W. G. (1978), 'The stochastic frontier production function and average efficiency: An empirical analysis', *Journal of Econometrics* **7**(3), 385 – 389.

Linton, O., Maasoumi, E. and Whang, Y.-J. (2005), 'Consistent Testing for Stochastic Dominance under General Sampling Schemes', *The Review of Economic Studies* **72**(3), 735–765.

McCloud, N. and Kumbhakar, S. C. (2008), 'Do subsidies drive productivity? a cross-country analysis of nordic dairy farms', *Advances in Econometrics* **23**(1), 245–274.

Mundlak, Y. (1961), 'Empirical Production Function Free of Management Bias', *American Journal of Agricultural Economics* **43**(1), 44–56.

Newey, W. K., Powell, J. L. and Vella, F. (1999), 'Nonparametric Estimation of Triangular Simultaneous Equations Models', *Econometrica* **67**(3), 565–603.

Parmeter, C. F. and Kumbhakar, S. C. (2014), 'Efficiency analysis: A primer on recent advances', *Foundations and Trends® in Econometrics* **7**(3–4), 191–385.

Politis, D. N. and Romano, J. P. (1994), 'Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions', *Ann. Statist.* **22**(4), 2031–2050.

Rho, S. and Schmidt, P. (2015), ‘Are all firms inefficient?’, *Journal of Productivity Analysis* **43**(3), 327–349.

Rotnitzky, A., Cox, D. R., Bottai, M. and Robins, J. (2000), ‘Likelihood-based inference with singular information matrix’, *Bernoulli* **6**(2), 243–284.

Sargan, J. D. (1983), ‘Identification and lack of identification’, *Econometrica* **51**(6), 1605–1633.

Schmidt, P. and Sickles, R. C. (1984), ‘Production frontiers and panel data’, *Journal of Business & Economic Statistics* **2**(4), 367–374.

Self, S. G. and Liang, K.-Y. (1987), ‘Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions’, *Journal of the American Statistical Association* **82**(398), 605–610.

Simar, L., Knox Lovell, C. and Vanden Eeckaut, P. (1994), ‘Stochastic frontiers incorporating exogenous influences on efficiency’, *STAT Discussion Papers* (9403).

Sundberg, R. (1974*a*), ‘Maximum likelihood theory for incomplete data from an exponential family’, *Scandinavian Journal of Statistics* **1**(2), 49–58.

Sundberg, R. (1974*b*), ‘On estimation and testing for the folded normal distribution’, *Communications in Statistics* **3**(1), 55–72.

Wasylenko, M. J. and Erickson, R. A. (1978), ‘“On Measuring Economic Diversification": Comment’, *Land Economics* **54**(1), 106–109.

Wilde, J. (2008), ‘A simple representation of the Bera–Jarque–Lee test for probit models’, *Economics Letters* **101**(2), 119 – 121.

Wooldridge, J. M. (2015), ‘Control Function Methods in Applied Econometrics’, *Journal of Human Resources* **50**(2), 420–445.

Appendix A. Technical derivations

A.1. **Derivation of the likelihood function.** We provide here the main steps for the calculation of the conditional distribution of η given ε. Recall that the joint density of (ε, η) is given by

$$
\begin{aligned}
f_{\varepsilon,\eta}(\varepsilon,\eta) = & \frac{1}{2\pi\tilde{\sigma}(Z)}\Bigg\{\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)} + \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)} - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\times \\
& \exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta+\rho_U\sigma_U(Z)\eta)^2}{2\tilde{\sigma}^2(Z)} - \frac{\eta^2}{2}\right) \\
& + \Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)} - \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)} - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\times \\
& \exp\left(-\frac{(\varepsilon-\rho_V\sigma_V\eta-\rho_U\sigma_U(Z)\eta)^2}{2\tilde{\sigma}^2(Z)} - \frac{\eta^2}{2}\right)\Bigg\}.
\end{aligned}
$$

We analyze the kernel of the two components of this distribution separately.

(i) Let $\rho_1(Z) = \rho_V\sigma_V - \rho_U\sigma_U(Z)$. The kernel of the first component is equal to

$$
\begin{aligned}
& \frac{1}{\tilde{\sigma}^2(Z)}\left(\varepsilon^2 - 2\rho_1(Z)\varepsilon\eta + (\tilde{\sigma}^2(Z)+\rho_1^2(Z))\eta^2\right) \\
& \quad = \frac{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}{2\tilde{\sigma}^2(Z)}\left(\eta^2 - 2\frac{\rho_1(Z)\varepsilon\eta}{\tilde{\sigma}^2(Z)+\rho_1^2(Z)} + \frac{\varepsilon^2}{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}\right) \\
& \quad = \frac{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}{2\tilde{\sigma}^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2 + \frac{\varepsilon^2}{2(\tilde{\sigma}^2(Z)+\rho_1^2(Z))}.
\end{aligned}
$$

(ii) Let $\rho_2(Z) = \rho_V\sigma_V + \rho_U\sigma_U(Z)$. The kernel of the second component can be similarly written as

$$
\begin{aligned}
& \frac{1}{2\tilde{\sigma}(Z)^2}\left(\varepsilon^2 - 2\rho_2(Z)\varepsilon\eta + (\tilde{\sigma}^2(Z)+\rho_2^2(Z))\eta^2\right) \\
& \quad = \frac{\tilde{\sigma}^2(Z)+\rho_2^2(Z)}{2\tilde{\sigma}^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\tilde{\sigma}^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2 + \frac{\varepsilon^2}{2(\tilde{\sigma}^2(Z)+\rho_2^2(Z))}.
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
f_{\varepsilon,\eta}(\varepsilon,\eta) = & \frac{1}{2\pi\tilde{\sigma}(Z)}\Bigg\{\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)} + \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)} - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\times \\
& \exp\left(-\frac{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}{2\tilde{\sigma}^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\tilde{\sigma}^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2 - \frac{\varepsilon^2}{2(\tilde{\sigma}^2(Z)+\rho_1^2(Z))}\right)
\end{aligned}
$$

$$+\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\tilde{\sigma}(Z)}-\frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\tilde{\sigma}(Z)}-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\times$$
$$\exp\left(-\frac{\tilde{\sigma}^2(Z)+\rho_2^2(Z)}{2\tilde{\sigma}^2(Z)}\left(\eta-\frac{\rho_2(Z)}{\tilde{\sigma}^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2-\frac{\varepsilon^2}{2(\tilde{\sigma}^2(Z)+\rho_2^2(Z))}\right)\Bigg\}.$$

We would like to integrate out the random variable η. To do so, we rearrange the terms above as follows

$$\begin{aligned}
f_{\varepsilon,\eta}(\varepsilon,\eta)=&\Bigg\{\Phi\left(q_1(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\frac{1}{\sqrt{2\pi\sigma^2_{\eta,1}(Z)}}\exp\left(-\frac{(\eta-\mu_{\eta,1}(Z)\varepsilon)^2}{2\sigma^2_{\eta,1}(Z)}\right)\times\\
&\frac{1}{\sqrt{2\pi\sigma^2_{\varepsilon,1}(Z)}}\exp\left(-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,1}(Z)}\right)\\
&+\Phi\left(q_2(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\frac{1}{\sqrt{2\pi\sigma^2_{\eta,2}(Z)}}\exp\left(-\frac{(\eta-\mu_{\eta,2}(Z)\varepsilon)^2}{2\sigma^2_{\eta,2}(Z)}\right)\times\\
&\frac{1}{\sqrt{2\pi\sigma^2_{\varepsilon,2}(Z)}}\exp\left(-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,2}(Z)}\right)\Bigg\}\\
=&\frac{1}{2}\Bigg\{\frac{\Phi\left(q_1(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)}{\Phi\left(\tau_1(Z)\varepsilon\right)}\frac{1}{\sqrt{2\pi\sigma^2_{\eta,1}(Z)}}\exp\left(-\frac{(\eta-\mu_{\eta,1}(Z)\varepsilon)^2}{2\sigma^2_{\eta,1}(Z)}\right)\times\\
&\frac{2}{\sqrt{2\pi\sigma^2_{\varepsilon,1}(Z)}}\Phi\left(\tau_1(Z)\varepsilon\right)\exp\left(-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,1}(Z)}\right)\\
&+\frac{\Phi\left(q_2(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)}{\Phi\left(\tau_2(Z)\varepsilon\right)}\frac{1}{\sqrt{2\pi\sigma^2_{\eta,2}(Z)}}\exp\left(-\frac{(\eta-\mu_{\eta,2}(Z)\varepsilon)^2}{2\sigma^2_{\eta,2}(Z)}\right)\times\\
&\frac{2}{\sqrt{2\pi\sigma^2_{\varepsilon,2}(Z)}}\Phi\left(\tau_2(Z)\varepsilon\right)\exp\left(-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,2}(Z)}\right)\Bigg\}\\
=&\frac{1}{2}\left\{f_{\varepsilon,\eta,1}\left(\varepsilon,\eta\right)+f_{\varepsilon,\eta,2}\left(\varepsilon,\eta\right)\right\},
\end{aligned}$$

where

$$\sigma^2_{\varepsilon,1}(Z) = \sigma^2_V + \sigma^2_U(Z) - 2\rho_V\sigma_V\rho_U\sigma_U(Z), \qquad \sigma^2_{\varepsilon,2}(Z) = \sigma^2_V + \sigma^2_U(Z) + 2\rho_V\sigma_V\rho_U\sigma_U(Z),$$

$$q_1(Z) = \frac{\lambda(Z)\rho_V\sigma_V}{\tilde{\sigma}(Z)} + \frac{\rho_U\sigma_U(Z)}{\lambda(Z)\tilde{\sigma}(Z)}, \qquad q_2(Z) = \frac{\lambda(Z)\rho_V\sigma_V}{\tilde{\sigma}(Z)} - \frac{\rho_U\sigma_U(Z)}{\lambda(Z)\tilde{\sigma}(Z)},$$

$$\mu_{\eta,1}(Z) = \frac{\rho_1(Z)}{\sigma^2_{\varepsilon,1}(Z)}, \qquad \mu_{\eta,2}(Z) = \frac{\rho_2(Z)}{\sigma^2_{\varepsilon,2}(Z)},$$

$$\sigma^2_{\eta,1}(Z) = \frac{\tilde{\sigma}^2(Z)}{\sigma^2_{\varepsilon,1}(Z)}, \qquad \sigma^2_{\eta,2}(Z) = \frac{\tilde{\sigma}^2(Z)}{\sigma^2_{\varepsilon,2}(Z)},$$

$$\tau_1(Z) = \frac{\mu_{\eta,1}(Z)q_1(Z) - \frac{\lambda(Z)}{\tilde{\sigma}(Z)}}{\sqrt{1 + q_1^2(Z)\sigma^2_{\eta,1}(Z)}}, \qquad \tau_2(Z) = \frac{\mu_{\eta,2}(Z)q_2(Z) - \frac{\lambda(Z)}{\tilde{\sigma}(Z)}}{\sqrt{1 + q_2^2(Z)\sigma^2_{\eta,2}(Z)}}.$$

and the normalizing factor $\tau_j(Z)$, with $j = 1, 2$ comes from the fact that

$$\Phi\left(\tau_j(Z)\varepsilon\right) = \int_{-\infty}^{\infty} \Phi\left(q_j(Z)\eta - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right) \frac{1}{\sqrt{2\pi\sigma^2_{\eta,j}(Z)}} \exp\left(-\frac{(\eta - \mu_{\eta,j}(Z)\varepsilon)^2}{2\sigma^2_{\eta,j}(Z)}\right) d\eta,$$

by Lemma 2.2. in Azzalini (2013, p. 26).

Each component of this density can be interpreted as the pdf of a bivariate skew-normal distribution, properly rearranged into the product of a conditional and a marginal density (Azzalini and Dalla Valle, 1996). Therefore, the full joint density is a mixture of two skew normal distribution with equal weights, 0.5. To obtain the conditional cdf of η given ε, we need to integrate the above expression appropriately. The following integral

$$\Psi_{0,j}(Z,\theta) = \int_{-\infty}^{-\tilde{W}\gamma} \Phi\left(q_j(Z)\eta - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right) \frac{1}{\sqrt{2\pi\sigma^2_{\eta,j}(Z)}} \exp\left(-\frac{(\eta - \mu_{\eta,j}(Z)\varepsilon)^2}{2\sigma^2_{\eta,j}(Z)}\right) d\eta,$$

cannot be directly evaluated analytically. However, using the properties of the skew-normal distribution, it can be expressed as the cdf of a bivariate normal distribution.

Let us define the fictitious random vector (η, κ_j), for $j = 1, 2$, such that the conditional distribution of (η, κ_j) given ε is a bivariate normal distribution. That is, we have

$$\begin{matrix} \frac{\eta - \mu_{\eta,j}(Z)\varepsilon}{\sigma_{\eta,j}(Z)} \\ \frac{\kappa_j + q_j(Z)\mu_{\eta,j}(Z)\varepsilon - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}}{\sqrt{1+q_j^2(Z)\sigma^2_{\eta,j}(Z)}}\varepsilon \end{matrix} \Bigg| \varepsilon \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_j^*(Z) \\ \rho_j^*(Z) & 1 \end{bmatrix}\right),$$

where,

$$\rho_j^*(Z) = -\frac{q_j(Z)\sigma_{\eta,j}(Z)}{\sqrt{1+q_j^2(Z)\sigma_{\eta,j}^2(Z)}}.$$

Let $\Phi_2(\cdot,\cdot;\rho^*(Z))$, be the cdf of a standard bivariate normal distribution with correlation coefficient equal to $\rho^*(Z)$. We have that

$$\begin{aligned}\int_{a_j}^{b_j} &\Phi\left(q_j(Z)\eta - \frac{\lambda(Z)\varepsilon}{\tilde{\sigma}(Z)}\right)\frac{1}{\sqrt{2\pi\sigma_{\eta,j}^2(Z)}}\exp\left(-\frac{(\eta-\mu_{\eta,j}(Z)\varepsilon)^2}{2\sigma_{\eta,j}^2(Z)}\right)d\eta \\ =&\Phi_2\left(\frac{b_j-\mu_{\eta,j}(Z)\varepsilon}{\sigma_{\eta,j}(Z)},\tau_j(Z)\varepsilon;\rho^*(Z)\right)-\Phi_2\left(\frac{a_j-\mu_{\eta,j}(Z)\varepsilon}{\sigma_{\eta,j}(Z)},\tau_j(Z)\varepsilon;\rho^*(Z)\right).\end{aligned}$$

This finally implies that

$$\Psi_{0,j}(Z,\theta) = \Phi_2\left(\frac{-\tilde{W}\gamma-\mu_{\eta,j}(Z)\varepsilon}{\sigma_{\eta,j}(Z)},\tau_j(Z)\varepsilon;\rho^*(Z)\right),$$

and

$$\Psi_{1,j}(Z,\theta) = \Phi\left(\tau_j(Z)\varepsilon\right)-\Phi_2\left(\frac{-\tilde{W}\gamma-\mu_{\eta,j}(Z)\varepsilon}{\sigma_{\eta,j}(Z)},\tau_j(Z)\varepsilon;\rho^*(Z)\right).$$

Ultimately, these integrals only involve a bivariate normal cumulative distribution function, which is readily available in any standard statistical software.

Therefore, the likelihood function is given by

$$\begin{aligned}\mathcal{L}(\theta) = &\left(\sum_{j=1,2}\Psi_{1,j}(Z,\theta)\frac{1}{\sigma_{\varepsilon,j}(Z)}\phi\left(\frac{\varepsilon}{\sigma_{\varepsilon,j}(Z)}\right)\right)^Z \times \\ &\left(\sum_{j=1,2}\Psi_{0,j}(Z,\theta)\frac{1}{\sigma_{\varepsilon,j}(Z)}\phi\left(\frac{\varepsilon}{\sigma_{\varepsilon,j}(Z)}\right)\right)^{1-Z},\end{aligned}$$

where $\phi(\cdot)$ denotes the pdf of a standard normal distribution.

A.2. **Proof of Proposition 2.1.** By rearranging terms, we can write the likelihood function as follows

$$\mathcal{L}(\theta) = \left(\Psi_{1,1}(Z,\theta)\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma_{\varepsilon,1}^2(Z)}}\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma_{\varepsilon,1}^2(Z)\sigma_{\varepsilon,2}^2(Z)}\right)+\Psi_{1,2}(Z,\theta)\right)^Z \times$$

$$\left(\Psi_{0,1}(Z,\theta)\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right)+\Psi_{0,2}(Z,\theta)\right)^{1-Z}\times$$
$$\frac{1}{\sigma_{\varepsilon,2}(Z)}\phi\left(\frac{\varepsilon}{\sigma_{\varepsilon,2}(Z)}\right),$$

in a way that the log-likelihood function is given by

$$\begin{aligned}
\ell_n(\theta) =& Z\log\left\{\Psi_{1,1}(Z,\theta)\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right)+\Psi_{1,2}(Z,\theta)\right\} \\
&+(1-Z)\log\left\{\Psi_{0,1}(Z,\theta)\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right)+\Psi_{0,2}(Z,\theta)\right\} \\
&-\frac{1}{2}\log(\sigma^2_{\varepsilon,2}(Z))-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,2}(Z)} \\
=& Z\log(D_1)+(1-Z)\log(D_0)-\frac{1}{2}\log(\sigma^2_{\varepsilon,2}(Z))-\frac{\varepsilon^2}{2\sigma^2_{\varepsilon,2}(Z)},
\end{aligned}$$

where the definition of the terms D_0 and D_1 should be apparent and we have omitted terms that are constant with respect to θ. Notice that, when $\rho_U = 0$, then $\Psi_j(Z,\theta) = \Psi_{j,1}(Z,\theta) = \Psi_{j,2}(Z,\theta)$, $D_j = 2\Psi_j(Z,\theta)$, and

$$\begin{aligned}
\frac{\partial\Psi_{j,1}(Z,\theta)}{\partial\rho_U} &= -\frac{\partial\Psi_{j,2}(Z,\theta)}{\partial\rho_U}, \\
\frac{\partial^2\Psi_{j,1}(Z,\theta)}{\partial\rho_U^2} &= \frac{\partial^2\Psi_{j,2}(Z,\theta)}{\partial\rho_U^2}
\end{aligned}$$

for $j = 0, 1$.

To prove Part (i) of the Proposition (ρ_U is only identified up to a sign), we use the fact that the log-likelihood function is an even function of ρ_U. That is, $\ell_n(\theta_1,\rho_U) = \ell_n(\theta_1,-\rho_U)$. Therefore, for a given θ_1, the first partial derivative of the log-likelihood wrt ρ_U satisfies

$$\frac{\partial\ell_n(\theta_1,\rho_U)}{\partial\rho_U} = -\frac{\partial\ell_n(\theta_1,-\rho_U)}{\partial\rho_U}.$$

At the true parameter value, we therefore have that

$$E\left[\frac{\partial\ell_n(\theta_{1,0},\rho_{U,0})}{\partial\rho_U}\right] = -E\left[\frac{\partial\ell_n(\theta_{1,0},-\rho_{U,0})}{\partial\rho_U}\right] = 0.$$

To prove Part (ii) (the derivative of the log-likelihood wrt ρ_U is identically equal to 0 for any value of θ_1), notice that by Rolle's theorem if $\ell_n(\theta_1, \rho_U) = \ell_n(\theta_1, -\rho_U)$, then there must exist a value of $r \in [-\rho_U, \rho_U]$, such that,

$$\frac{\partial \ell_n(\theta_1, r)}{\partial \rho_U} = 0.$$

In particular, since the log-likelihood is continuous and symmetric about $r = 0$ as a function of ρ_U, using a limiting argument, we must have that $\frac{\partial \ell_n(\theta_1, 0)}{\partial \rho_U} = 0$ for any θ_1.

We prove the result about the second derivative wrt ρ_U using direct computations. In the following, we let $SN(\xi, \omega^2, \alpha, \tau)$ be an extended skew-normal random variable with location parameter ξ, scale parameter ω, slant parameter α, and extended parameter τ (see Azzalini, 2013, p. 36-37). When $\tau = 0$, this random variable reduces to a skew-normal. In that case, the last parameter is simply omitted.

To simplify notations let

$$\zeta = \sqrt{1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}} \exp\left(-\frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right). \tag{9}$$

Thus we have that

$$\begin{aligned}
\frac{\partial \zeta}{\partial \rho_U} =& \frac{1}{\sqrt{1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}} \left(\frac{2\rho_V \sigma_V \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)} + \frac{4\rho_V^2 \sigma_V^2 \rho_U \sigma_U^2(Z)}{\sigma^4_{\varepsilon,1}(Z)}\right) \exp\left(-\frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right) \\
&- \sqrt{1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}} \left(\frac{2\varepsilon^2 \rho_V \sigma_V \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)} - \frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^4_{\varepsilon,1}(Z)\sigma^4_{\varepsilon,2}(Z)} \frac{\partial \sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial \rho_U}\right) \times \\
&\quad \exp\left(-\frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \zeta}{\partial \rho_U^2} =& -\frac{1}{4\left(1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}\right)^{\frac{3}{2}}} \left(\frac{4\rho_V \sigma_V \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)} + \frac{8\rho_V^2 \sigma_V^2 \rho_U \sigma_U^2(Z)}{\sigma^4_{\varepsilon,1}(Z)}\right)^2 \exp\left(-\frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right) \\
&+ \frac{1}{2\sqrt{1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}} \left(\frac{16\rho_V^2 \sigma_V^2 \sigma_U^2(Z)}{\sigma^4_{\varepsilon,1}(Z)} + \frac{32\rho_V^3 \sigma_V^3 \rho_U \sigma_U^3(Z)}{\sigma^6_{\varepsilon,1}(Z)}\right) \exp\left(-\frac{2\varepsilon^2 \rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right) \\
&- \frac{1}{\sqrt{1 + \frac{4\rho_V \sigma_V \rho_U \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}} \left(\frac{4\rho_V \sigma_V \sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)} + \frac{8\rho_V^2 \sigma_V^2 \rho_U \sigma_U^2(Z)}{\sigma^4_{\varepsilon,1}(Z)}\right) \times
\end{aligned}$$

$$\left(\frac{2\varepsilon^2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)} - \frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^4_{\varepsilon,1}(Z)\sigma^4_{\varepsilon,2}(Z)}\frac{\partial\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial\rho_U}\right)\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right)$$
$$-\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}\left(-\frac{4\varepsilon^2\rho_V\sigma_V\sigma_U(Z)}{\sigma^4_{\varepsilon,1}(Z)\sigma^4_{\varepsilon,2}(Z)}\frac{\partial\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial\rho_U} - \frac{4\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^8_{\varepsilon,1}(Z)\sigma^8_{\varepsilon,2}(Z)}\frac{\partial^2\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial\rho_U^2}\right)\times$$
$$\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right)$$
$$+\sqrt{1+\frac{4\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)}}\left(\frac{2\varepsilon^2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)} - \frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^4_{\varepsilon,1}(Z)\sigma^4_{\varepsilon,2}(Z)}\frac{\partial\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial\rho_U}\right)^2\times$$
$$\exp\left(-\frac{2\varepsilon^2\rho_V\sigma_V\rho_U\sigma_U(Z)}{\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}\right).$$

When $\rho_U = 0$, then $\zeta = 1$, and

$$\left.\frac{\partial\zeta}{\partial\rho_U}\right|_{\rho_U=0} = \frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right),$$
$$\left.\frac{\partial^2\zeta}{\partial\rho_U^2}\right|_{\rho_U=0} = \frac{4\rho_V^2\sigma_V^2\sigma_U^2(Z)}{\sigma^4(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)^2,$$

where the last expression follows from the fact that

$$\left.\frac{\partial\sigma^2_{\varepsilon,1}(Z)\sigma^2_{\varepsilon,2}(Z)}{\partial\rho_U}\right|_{\rho_U=0} = \left(-2\rho_V\sigma_V\sigma_U(Z)\sigma^2_{\varepsilon,2}(Z) + 2\rho_V\sigma_V\sigma_U(Z)\sigma^2_{\varepsilon,1}(Z)\right)_{\rho_U=0} = 0.$$

Therefore, we finally have that

$$\left(\left.\frac{\partial\zeta}{\partial\rho_U}\right|_{\rho_U=0}\right)^2 = \left.\frac{\partial^2\zeta}{\partial\rho_U^2}\right|_{\rho_U=0}.$$

The first derivative of the log-likelihood can be written as

$$\begin{aligned}\frac{\partial\ell_n(\theta)}{\partial\rho_U} =& \frac{Z}{D_1}\left[\frac{\partial\Psi_{1,1}(Z,\theta)}{\partial\rho_U}\zeta + \Psi_{1,1}(Z,\theta)\frac{\partial\zeta}{\partial\rho_U} + \frac{\partial\Psi_{1,2}(Z,\theta)}{\partial\rho_U}\right]\\ &+ \frac{1-Z}{D_0}\left[\frac{\partial\Psi_{0,1}(Z,\theta)}{\partial\rho_U}\zeta + \Psi_{0,1}(Z,\theta)\frac{\partial\zeta}{\partial\rho_U} + \frac{\partial\Psi_{0,2}(Z,\theta)}{\partial\rho_U}\right]\\ &- \frac{\sigma_V\rho_V\sigma_U(Z)}{\sigma^2_{\varepsilon,2}(Z)} + \frac{\varepsilon^2\sigma_V\rho_V\sigma_U(Z)}{\sigma^4_{\varepsilon,2}(Z)}.\end{aligned}$$

Thus, when $\rho_U = 0$,

$$\begin{aligned}\left.\frac{\partial\ell_n(\theta)}{\partial\rho_U}\right|_{\rho_U=0} =& \frac{Z}{2\Psi_1(Z,\theta)}\left[\Psi_1(Z,\theta)\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]\\ &+ \frac{1-Z}{2\Psi_0(Z,\theta)}\left[\Psi_0(Z,\theta)\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]\end{aligned}$$

$$
- \frac{\sigma_V \rho_V \sigma_U(Z)}{\sigma^2(Z)} + \frac{\varepsilon^2 \sigma_V \rho_V \sigma_U(Z)}{\sigma^4(Z)}
$$

$$
= \frac{\rho_V \sigma_V \sigma_U(Z)}{\sigma^2(Z)} \left(1 - \frac{\varepsilon^2}{\sigma^2(Z)}\right) - \frac{\sigma_V \rho_V \sigma_U(Z)}{\sigma^2(Z)} + \frac{\varepsilon^2 \sigma_V \rho_V \sigma_U(Z)}{\sigma^4(Z)} = 0,
$$

which confirms the result shown above.

The second derivative of the log-likelihood is

$$
\begin{aligned}
\frac{\partial^2 \ell_n(\theta)}{\partial \rho_U^2} = & - \frac{Z}{D_1^2} \left[\frac{\partial \Psi_{1,1}(Z,\theta)}{\partial \rho_U} \zeta + \Psi_{1,1}(Z,\theta) \frac{\partial \zeta}{\partial \rho_U} + \frac{\partial \Psi_{1,2}(Z,\theta)}{\partial \rho_U} \right]^2 \\
& + \frac{Z}{D_1} \left[\frac{\partial^2 \Psi_{1,1}(Z,\theta)}{\partial \rho_U^2} \zeta + 2 \frac{\partial \Psi_{1,1}(Z,\theta)}{\partial \rho_U} \frac{\partial \zeta}{\partial \rho_U} + \Psi_{1,1}(Z,\theta) \frac{\partial^2 \zeta}{\partial \rho_U^2} + \frac{\partial^2 \Psi_{1,2}(Z,\theta)}{\partial \rho_U^2} \right] \\
& - \frac{1-Z}{D_0^2} \left[\frac{\partial \Psi_{0,1}(Z,\theta)}{\partial \rho_U} + \Psi_{0,1}(Z,\theta) \frac{\partial \zeta}{\partial \rho_U} + \frac{\partial \Psi_{0,2}(Z,\theta)}{\partial \rho_U} \right]^2 \\
& + \frac{1-Z}{D_0} \left[\frac{\partial^2 \Psi_{0,1}(Z,\theta)}{\partial \rho_U^2} \zeta + 2 \frac{\partial \Psi_{0,1}(Z,\theta)}{\partial \rho_U} \frac{\partial \zeta}{\partial \rho_U} + \Psi_{0,1}(Z,\theta) \frac{\partial^2 \zeta}{\partial \rho_U^2} + \frac{\partial^2 \Psi_{0,2}(Z,\theta)}{\partial \rho_U^2} \right] \\
& + \frac{2\sigma_V^2 \rho_V^2 \sigma_U^2(Z)}{\sigma_{\varepsilon,2}^4(Z)} - \frac{4\varepsilon^2 \sigma_V^2 \rho_V^2 \sigma_U^2(Z)}{\sigma_{\varepsilon,2}^6(Z)}.
\end{aligned}
$$

Thus, when $\rho_U = 0$,

$$
\begin{aligned}
\left. \frac{\partial^2 \ell_n(\theta)}{\partial \rho_U^2} \right|_{\rho_U=0} = & - \frac{Z}{4\Psi_1^2(Z,\theta)} \left[\Psi_1(Z,\theta) \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} \right]^2 \\
& + \frac{Z}{2\Psi_1(Z,\theta)} \left[2 \frac{\partial^2 \Psi_1(Z,\theta)}{\partial \rho_U^2} + 2 \frac{\partial \Psi_1(Z,\theta)}{\partial \rho_U} \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} + \Psi_1(Z,\theta) \left. \frac{\partial^2 \zeta}{\partial \rho_U^2} \right|_{\rho_U=0} \right] \\
& - \frac{1-Z}{4\Psi_0^2(Z,\theta)} \left[\Psi_0(Z,\theta) \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} \right]^2 \\
& + \frac{1-Z}{2\Psi_0(Z,\theta)} \left[2 \frac{\partial^2 \Psi_0(Z,\theta)}{\partial \rho_U^2} + 2 \frac{\partial \Psi_0(Z,\theta)}{\partial \rho_U} \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} + \Psi_0(Z,\theta) \left. \frac{\partial^2 \zeta}{\partial \rho_U^2} \right|_{\rho_U=0} \right] \\
& + \frac{2\sigma_V^2 \rho_V^2 \sigma_U^2(Z)}{\sigma^4(Z)} - \frac{4\varepsilon^2 \sigma_V^2 \rho_V^2 \sigma_U^2(Z)}{\sigma^6(Z)} \\
= & \frac{1}{4} \left(\left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} \right)^2 + \frac{Z}{\Psi_1(Z,\theta)} \left[\frac{\partial^2 \Psi_1(Z,\theta)}{\partial \rho_U^2} + \frac{\partial \Psi_1(Z,\theta)}{\partial \rho_U} \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} \right] \\
& + \frac{1-Z}{\Psi_0(Z,\theta)} \left[\frac{\partial^2 \Psi_0(Z,\theta)}{\partial \rho_U^2} + \frac{\partial \Psi_0(Z,\theta)}{\partial \rho_U} \left. \frac{\partial \zeta}{\partial \rho_U} \right|_{\rho_U=0} \right]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2\sigma_V^2\rho_V^2\sigma_U^2(Z)}{\sigma^4(Z)}-\frac{4\varepsilon^2\sigma_V^2\rho_V^2\sigma_U^2(Z)}{\sigma^6(Z)}\\
=&\frac{\rho_V^2\sigma_V^2\sigma_U^2(Z)}{\sigma^4(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)^2+\frac{2\sigma_V^2\rho_V^2\sigma_U^2(Z)}{\sigma^4(Z)}-\frac{4\varepsilon^2\sigma_V^2\rho_V^2\sigma_U^2(Z)}{\sigma^6(Z)}\\
&+\frac{Z}{\Psi_1(Z,\theta)}\left[\frac{\partial^2\Psi_1(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_1(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]\\
&+\frac{1-Z}{\Psi_0(Z,\theta)}\left[\frac{\partial^2\Psi_0(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_0(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]\\
=&\frac{\rho_V^2\sigma_V^2\sigma_U^2(Z)}{\sigma^4(Z)}\left(3-\frac{6\varepsilon^2}{\sigma^2(Z)}+\frac{\varepsilon^4}{\sigma^4(Z)}\right)\\
&+\frac{Z}{\Psi_1(Z,\theta)}\left[\frac{\partial^2\Psi_1(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_1(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]\\
&+\frac{1-Z}{\Psi_0(Z,\theta)}\left[\frac{\partial^2\Psi_0(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_0(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right]
\end{aligned}
$$

We take the expectation wrt the joint density of $(\varepsilon, Z, \tilde{W})$. Because $(\varepsilon,\eta) \perp\!\!\!\perp \tilde{W}$ and $\varepsilon \perp\!\!\!\perp Z|\eta$, then

$$
\begin{aligned}
f_{\varepsilon,Z,\tilde{W}}(e,z,\tilde{w}) =& f_{\varepsilon|Z,\tilde{W}}(e|z,\tilde{w})\,P(Z=z|\tilde{w})\,f_{\tilde{W}}(\tilde{w})\\
=&\begin{cases}
f_{\varepsilon|\eta\le -,\tilde{W}}(e|\eta\le -\tilde{w}\gamma,\tilde{w})\,\Phi(-\tilde{w}\gamma)\,f_{\tilde{W}}(\tilde{w}) & \\
\qquad = F_{\eta|\varepsilon}(-\tilde{w}\gamma|e)\,f_\varepsilon(e)f_{\tilde{W}}(\tilde{w}) & \text{if } z=0\\
f_{\varepsilon|\eta\ge -\tilde{W}\gamma,\tilde{W}}(e|\eta\ge -\tilde{w}\gamma,\tilde{w})\,(1-\Phi(-\tilde{w}\gamma))\,f_{\tilde{W}}(\tilde{w}) & \\
\qquad = \left[1-F_{\eta|\varepsilon}(-\tilde{w}\gamma|e)\right]f_\varepsilon(e)f_{\tilde{W}}(\tilde{w}) & \text{if } z=1
\end{cases}
\end{aligned}
$$

Therefore, for any function $g(\varepsilon)$,

$$
E[Zg(\varepsilon)] = \int_{\tilde{w}}\left(\int_\varepsilon \frac{\Psi_1(1,\theta)}{\Phi(\tau(1)\varepsilon)}g(\varepsilon)f_{\varepsilon,1}(\varepsilon)d\varepsilon\right)f_{\tilde{W}}(\tilde{w})d\tilde{w} = E_{\tilde{W}}\left[\int_\varepsilon \frac{\Psi_1(1,\theta)}{\Phi(\tau(1)\varepsilon)}g(\varepsilon)f_{\varepsilon,1}(\varepsilon)d\varepsilon\right],
$$

and

$$
E[(1-Z)g(\varepsilon)] = \int_{\tilde{w}}\left(\int_\varepsilon \frac{\Psi_0(0,\theta)}{\Phi(\tau(0)\varepsilon)}f_{\varepsilon,0}(\varepsilon)d\varepsilon\right)f_{\tilde{W}}(\tilde{w})d\tilde{w} = E_{\tilde{W}}\left[\int_\varepsilon \frac{\Psi_0(0,\theta)}{\Phi(\tau(0)\varepsilon)}g(\varepsilon)f_{\varepsilon,0}(\varepsilon)d\varepsilon\right],
$$

with

$$
f_{\varepsilon,z}(\varepsilon) = \frac{2\Phi(\tau(z)\varepsilon)}{\sigma(z)}\phi\left(\frac{\varepsilon}{\sigma(z)}\right), \quad z\in\{0,1\}.
$$

Therefore,

$$
\begin{aligned}
E\left[\frac{\partial^2 \ell_n(\theta)}{\partial \rho_U^2}\bigg|_{\rho_U=0}\right] &= E\left[\frac{\rho_V^2\sigma_V^2\sigma_U^2(Z)}{\sigma^4(Z)}\left(3-\frac{6\varepsilon^2}{\sigma^2(Z)}+\frac{\varepsilon^4}{\sigma^4(Z)}\right)\right]\\
&+E\left[\frac{Z}{\Psi_1(Z,\theta)}\left(\frac{\partial^2\Psi_1(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_1(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right)\right]\\
&+E\left[\frac{1-Z}{\Psi_0(Z,\theta)}\left(\frac{\partial^2\Psi_0(Z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_0(Z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(Z)}{\sigma^2(Z)}\left(1-\frac{\varepsilon^2}{\sigma^2(Z)}\right)\right)\right]\\
=&E\left[\frac{\rho_V^2\sigma_V^2\sigma_U^2(Z)}{\sigma^4(Z)}\left(3-\frac{6E\left[\varepsilon^2|\eta\right]}{\sigma^2(Z)}+\frac{E\left[\varepsilon^4|\eta\right]}{\sigma^4(Z)}\right)\right] && (SD_1)\\
&+E_{\tilde W}\left[\int_\varepsilon\left(\frac{\partial^2\Psi_1(1,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_1(1,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(1)}{\sigma^2(1)}\left(1-\frac{\varepsilon^2}{\sigma^2(1)}\right)\right)\frac{2}{\sigma(1)}\phi\left(\frac{\varepsilon}{\sigma(1)}\right)d\varepsilon\right] && (SD_2)\\
&+E_{\tilde W}\left[\int_\varepsilon\left(\frac{\partial^2\Psi_0(0,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_0(0,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(0)}{\sigma^2(0)}\left(1-\frac{\varepsilon^2}{\sigma^2(0)}\right)\right)\frac{2}{\sigma(0)}\phi\left(\frac{\varepsilon}{\sigma(0)}\right)d\varepsilon\right]. && (SD_3)
\end{aligned}
$$

To treat the term is SD_1, when ρ_U is equal to zero, we notice that the density of ε conditional on η is a $SN(\rho_V\sigma_V\eta, \tilde\sigma(z), \lambda(z))$, which implies that

$$
\begin{aligned}
E\left[\varepsilon^2|\eta\right] =&\tilde\sigma^2(Z)+\sigma_V^2\rho_V^2\eta^2-2\sqrt{\frac{2}{\pi}}\sigma_U(Z)\sigma_V\rho_V\eta\\
=&\sigma^2(Z)+\sigma_V^2\rho_V^2\left(\eta^2-1\right)-2\sqrt{\frac{2}{\pi}}\sigma_U(Z)\sigma_V\rho_V\eta,
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\varepsilon^4|\eta\right] =&3\tilde\sigma^4(Z)-\left(12\sqrt{\frac{2}{\pi}}\tilde\sigma^2(Z)\sigma_U(Z)-4\sqrt{\frac{2}{\pi}}\sigma_U^3(Z)\right)\sigma_V\rho_V\eta+6\tilde\sigma^2(Z)\sigma_V^2\rho_V^2\eta^2\\
&-4\sqrt{\frac{2}{\pi}}\sigma_U(Z)\sigma_V^3\rho_V^3\eta^3+\sigma_V^4\rho_V^4\eta^4\\
=&3\sigma^4(Z)+3\sigma_V^4\rho_V^4-6\sigma^2(Z)\sigma_V^2\rho_V^2\\
&-\sqrt{\frac{2}{\pi}}\left(12\tilde\sigma_V^2+8\sigma_U^2(Z)\right)\sigma_V\rho_V\eta\sigma_U(z)\\
&+6\sigma^2(Z)\sigma_V^2\rho_V^2\eta^2-6\sigma_V^4\rho_V^4\eta^2-4\sqrt{\frac{2}{\pi}}\sigma_U(Z)\sigma_V^3\rho_V^3\eta^3+\sigma_V^4\rho_V^4\eta^4\\
=&3\sigma^4(Z)+\rho_V^4\sigma_V^4\left(3-6\eta^2+\eta^4\right)+6\sigma^2(Z)\sigma_V^2\rho_V^2\left(\eta^2-1\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \sqrt{\frac{2}{\pi}} \left(12\tilde{\sigma}_V^2 + 8\sigma_U^2(Z)\right) \sigma_V \rho_V \eta \sigma_U(z) \\
&- 4\sqrt{\frac{2}{\pi}} \sigma_U(Z) \sigma_V^3 \rho_V^3 \eta^3.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
3 - \frac{6E\left[\varepsilon^2|\eta\right]}{\sigma^2(Z)} + \frac{E\left[\varepsilon^4|\eta\right]}{\sigma^4(Z)} =& 3 - 6 - 6\frac{\sigma_V^2\rho_V^2\left(\eta^2-1\right)}{\sigma^2(Z)} + 12\sqrt{\frac{2}{\pi}}\frac{\sigma_U(Z)\sigma_V\rho_V\eta}{\sigma^2(Z)} \\
&+ 3 + \frac{\rho_V^4\sigma_V^4}{\sigma^4(z)}\left(3 - 6\eta^2 + \eta^4\right) + 6\frac{\sigma_V^2\rho_V^2\left(\eta^2-1\right)}{\sigma^2(z)} \\
&- \sqrt{\frac{2}{\pi}}\left(12\tilde{\sigma}_V^2 + 8\sigma_U^2(Z)\right)\frac{\sigma_U(z)\rho_V\sigma_V\eta}{\sigma^4(z)} - 4\sqrt{\frac{2}{\pi}}\frac{\sigma_U(Z)\sigma_V^3\rho_V^3\eta^3}{\sigma^4(z)} \\
=& \frac{\sigma_V^4\rho_V^4}{\sigma^4(Z)}\left[3 - 6\eta^2 + \eta^4\right] + 12\sqrt{\frac{2}{\pi}}\frac{\sigma_U(Z)\sigma_V^3\rho_V^3\eta}{\sigma^4(Z)} \\
&+ 4\sqrt{\frac{2}{\pi}}\frac{\sigma_U^3(z)\rho_V\sigma_V\eta}{\sigma^4(z)} - 4\sqrt{\frac{2}{\pi}}\frac{\sigma_U(Z)\sigma_V^3\rho_V^3\eta^3}{\sigma^4(z)}.
\end{aligned}
$$

When $Z = 0$, the distribution of $\eta|\eta \leq -\tilde{W}\gamma$ is a normal truncated above at $-\tilde{W}\gamma$. When $Z = 1$, the distribution of $\eta|\eta \geq -\tilde{W}\gamma$ is instead a normal truncated below at $-\tilde{W}\gamma$. Therefore, when $Z = 0$,

$$
\begin{aligned}
E\left[\eta|\eta \leq -\tilde{W}\gamma\right] =& -\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)} \\
E\left[\eta^2|\eta \leq -\tilde{W}\gamma\right] =& 1 + \tilde{W}\gamma\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)} \\
E\left[\eta^3|\eta \leq -\tilde{W}\gamma\right] =& -2\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)} - (\tilde{W}\gamma)^2\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)} \\
E\left[\eta^4|\eta \leq -\tilde{W}\gamma\right] =& 3 + 3\tilde{W}\gamma\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)} + (\tilde{W}\gamma)^3\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)},
\end{aligned}
$$

and when $Z = 1$

$$
\begin{aligned}
E\left[\eta|\eta \geq -\tilde{W}\gamma\right] =& \frac{\phi(-\tilde{W}\gamma)}{1 - \Phi(-\tilde{W}\gamma)} \\
E\left[\eta^2|\eta \geq -\tilde{W}\gamma\right] =& 1 - \tilde{W}\gamma\frac{\phi(-\tilde{W}\gamma)}{1 - \Phi(-\tilde{W}\gamma)} \\
E\left[\eta^3|\eta \geq -\tilde{W}\gamma\right] =& 2\frac{\phi(-\tilde{W}\gamma)}{1 - \Phi(-\tilde{W}\gamma)} + (\tilde{W}\gamma)^2\frac{\phi(-\tilde{W}\gamma)}{1 - \Phi(-\tilde{W}\gamma)}
\end{aligned}
$$

$$E\left[\eta^4 | \eta \geq -\tilde{W}\gamma\right] = 3 - 3\tilde{W}\gamma \frac{\phi(-\tilde{W}\gamma)}{1-\Phi(-\tilde{W}\gamma)} - (\tilde{W}\gamma)^3 \frac{\phi(-\tilde{W}\gamma)}{1-\Phi(-\tilde{W}\gamma)}.$$

Thus, we finally have that,

$$\begin{aligned} SD_1 = & \left(\frac{\rho_V^6\sigma_V^6\sigma_U^2(1)}{\sigma^8(1)} - \frac{\rho_V^6\sigma_V^6\sigma_U^2(0)}{\sigma^8(0)}\right) E_{\tilde{W}}\left[\left(3\tilde{W}\gamma - (\tilde{W}\gamma)^3\right)\phi(-\tilde{W}\gamma)\right] \\ & - 4\sqrt{\frac{2}{\pi}}\left(\frac{\sigma_U^3(1)\rho_V^5\sigma_V^5}{\sigma^8(1)} - \frac{\sigma_U^3(0)\rho_V^5\sigma_V^5}{\sigma^8(0)}\right) E_{\tilde{W}}\left[\left((\tilde{W}\gamma)^2 - 1\right)\phi(-\tilde{W}\gamma)\right] \\ & + 4\sqrt{\frac{2}{\pi}}\left(\frac{\sigma_U^5(1)\rho_V^3\sigma_V^3}{\sigma^8(1)} - \frac{\sigma_U^5(0)\rho_V^3\sigma_V^3}{\sigma^8(0)}\right) E_{\tilde{W}}\left[\phi(-\tilde{W}\gamma)\right] \end{aligned}$$

The first derivative of $\Psi_0(z,\theta)$ wrt ρ_U when $\rho_U = 0$ is equal to

$$\begin{aligned} \left.\frac{\partial \Psi_j(z,\theta)}{\partial \rho_U}\right|_{\rho_U=0} = & \Phi\left(-\lambda(z)\left(\frac{\varepsilon + \tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right)\right)\phi\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma + \frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right)\right)\times \\ & \left[\frac{\sigma_U(z)\tilde{\sigma}(z)}{\sigma^3(z)}\varepsilon + \frac{\rho_V\sigma_V\sigma_U(z)}{\sigma(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right] \\ & + \Phi\left(\frac{\tilde{W}\gamma - \frac{\rho_V}{\sigma_V}\varepsilon}{\sqrt{1-\rho_V^2}}\right)\phi\left(-\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right)\frac{\rho_V\sigma_V^2}{\sigma^3(z)}\varepsilon \\ & - \frac{\tilde{\sigma}_V^2}{\sigma(z)}\frac{1}{\tilde{\sigma}_V}\phi_{2,\rho^*(z)}\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma + \frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right), -\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right), \end{aligned}$$

where $\phi_{2,\rho^*(z)}$ is the kernel of the standard bivariate normal density with correlation parameter equal to $\rho^*(z)$.

Its second derivative wrt to ρ_U is instead equal to

$$\begin{aligned} \left.\frac{\partial^2 \Psi_0(z,\theta)}{\partial \rho_U^2}\right|_{\rho_U=0} = & \Phi\left(-\lambda(z)\left(\frac{\varepsilon + \tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right)\right)\phi\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma + \frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right)\right)\times \\ & \left\{\frac{\rho_V\sigma_V\sigma_U^2(z)}{\tilde{\sigma}^3(z)\sigma^5(z)}\left[3\rho_V^4\sigma_V^4 - 5\rho_V^2\sigma_V^2\sigma^2(z) + \sigma^4(z)\right]\varepsilon - \frac{\sigma_U^2(z)}{\tilde{\sigma}^3(z)\sigma^3(z)}\left[\tilde{\sigma}^4(z) + \rho_V^2\sigma_V^2\sigma_U^2(z)\right]\tilde{W}\gamma \right. \\ & \left. + \frac{\sigma(z)}{\tilde{\sigma}(z)}\left[\tilde{W}\gamma + \frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right]\left[\frac{\sigma_U(z)\tilde{\sigma}(z)}{\sigma^3(z)}\varepsilon + \frac{\rho_V\sigma_V\sigma_U(z)}{\sigma(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right]^2\right\} \\ & + \Phi\left(\frac{\tilde{W}\gamma - \frac{\rho_V}{\sigma_V}\varepsilon}{\sqrt{1-\rho_V^2}}\right)\phi\left(-\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right)\left[-\frac{\rho_V^2\sigma_U(z)\left(\sigma_U^4(z) + 3\sigma_U^2(z)\sigma_V^2 - \sigma_V^4\right)\varepsilon}{\sigma_V\sigma^5(z)} + \frac{\rho_V^2\sigma_V^3\sigma_U(z)}{\sigma^7(z)}\varepsilon^3\right] \\ & + \frac{\tilde{\sigma}(z)\sigma_V}{\sigma(z)}\frac{1}{\tilde{\sigma}_V}\phi_{2,\rho^*(z)}\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma + \frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right), -\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right)\times \end{aligned}$$

$$\left\{\frac{\rho_V(1-\rho_V^2)\sigma_U(z)\sigma^2(z)}{\tilde{\sigma}^3(z)} + \frac{\rho_V^3(1-\rho_V^2)\sigma_V^2\sigma_U^3(z)}{\tilde{\sigma}^3(z)\sigma^2(z)}\right.$$

$$\left. - \varepsilon^2\frac{\rho_V(1-\rho_V^2)\sigma_V^2\sigma_U(z)}{\tilde{\sigma}(z)\sigma^4(z)} - \tilde{W}\gamma\varepsilon\frac{(1-\rho_V^2)\sigma_V\sigma_U(z)}{\tilde{\sigma}(z)\sigma^2(z)} - (\tilde{W}\gamma)^2\frac{\rho_V(1-\rho_V^2)\sigma_V^2\sigma_U(z)}{\tilde{\sigma}^3(z)}\right\},$$

Thus

$$\int_\varepsilon \frac{\partial^2\Psi_0(z,\theta)}{\partial\rho_U^2} + \frac{\partial\Psi_0(z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(z)}{\sigma^2(z)}\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)d\varepsilon$$

$$= \int_\varepsilon \Phi\left(-\lambda(z)\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right)\right)\phi\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma+\frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right)\right)\times$$

$$\left\{2\left[\frac{\rho_V\sigma_V\sigma_U^2(z)\tilde{\sigma}(z)}{\sigma^5(z)}\varepsilon + \frac{\rho_V^2\sigma_V^2\sigma_U^2(z)}{\sigma^3(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right]\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)\right.$$

$$+ \frac{\rho_V\sigma_V\sigma_U^2(z)}{\tilde{\sigma}^3(z)\sigma^5(z)}\left[3\rho_V^4\sigma_V^4 - 5\rho_V^2\sigma_V^2\sigma^2(z)+\sigma^4(z)\right]\varepsilon - \frac{\sigma_U^2(z)}{\tilde{\sigma}^3(z)\sigma^3(z)}\left[\tilde{\sigma}^4(z)+\rho_V^2\sigma_V^2\sigma_U^2(z)\right]\tilde{W}\gamma$$

$$\left. + \frac{\sigma(z)}{\tilde{\sigma}(z)}\left[\tilde{W}\gamma+\frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right]\left[\frac{\sigma_U(z)\tilde{\sigma}(z)}{\sigma^3(z)}\varepsilon+\frac{\rho_V\sigma_V\sigma_U(z)}{\sigma(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right]^2\right\}\frac{2}{\sigma(z)}\phi\left(\frac{\varepsilon}{\sigma(z)}\right)d\varepsilon$$

$$+ \int_\varepsilon \Phi\left(\frac{\tilde{W}\gamma-\frac{\rho_V}{\sigma_V}\varepsilon}{\sqrt{1-\rho_V^2}}\right)\phi\left(-\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right)\left\{2\frac{\rho_V^2\sigma_V^3\sigma_U(z)}{\sigma^5(z)}\left(\varepsilon-\frac{\varepsilon^3}{\sigma^2(z)}\right)\right.$$

$$\left. -\frac{\rho_V^2\sigma_U(z)\left(\sigma_U^4(z)+3\sigma_U^2(z)\sigma_V^2-\sigma_V^4\right)\varepsilon}{\sigma_V\sigma^5(z)} + \frac{\rho_V^2\sigma_V^3\sigma_U(z)}{\sigma^7(z)}\varepsilon^3\right\}\frac{2}{\sigma(z)}\phi\left(\frac{\varepsilon}{\sigma(z)}\right)d\varepsilon$$

$$+ \int_\varepsilon \frac{1}{\tilde{\sigma}_V}\phi_{2,\rho^*(z)}\left(-\frac{\sigma(z)}{\tilde{\sigma}(z)}\left(\tilde{W}\gamma+\frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right), -\frac{\sigma_U(z)}{\sigma_V\sigma(z)}\varepsilon\right)\frac{2}{\sigma(z)}\phi\left(\frac{\varepsilon}{\sigma(z)}\right)$$

$$\left\{-2\frac{\rho_V\sigma_V\tilde{\sigma}_V^2\sigma_U(z)}{\sigma^3(z)}\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right) + \frac{\rho_V\sigma_V(1-\rho_V^2)\sigma_U(z)\sigma(z)}{\tilde{\sigma}^2(z)} + \frac{\rho_V^3(1-\rho_V^2)\sigma_V^4\sigma_U^3(z)}{\tilde{\sigma}(z)\sigma^4(z)}\right.$$

$$\left. - \varepsilon^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\sigma^5(z)} - \tilde{W}\gamma\varepsilon\frac{(1-\rho_V^2)\sigma_V^2\sigma_U(z)}{\sigma^3(z)} - (\tilde{W}\gamma)^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\tilde{\sigma}^2(z)\sigma(z)}\right\}d\varepsilon$$

$$=\phi(-\tilde{W}\gamma)\int_\varepsilon\left\{2\left[\frac{\rho_V\sigma_V\sigma_U^2(z)\tilde{\sigma}^2(z)}{\sigma^6(z)}\varepsilon + \frac{\rho_V^2\sigma_V^2\sigma_U^2(z)}{\sigma^4(z)}\tilde{W}\gamma\right]\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)\right.$$

$$+ \frac{\rho_V\sigma_V\sigma_U^2(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}\left[3\rho_V^4\sigma_V^4-5\rho_V^2\sigma_V^2\sigma^2(z)+\sigma^4(z)\right]\varepsilon - \frac{\sigma_U^2(z)}{\tilde{\sigma}^2(z)\sigma^4(z)}\left[\tilde{\sigma}^4(z)+\rho_V^2\sigma_V^2\sigma_U^2(z)\right]\tilde{W}\gamma$$

$$\left. + \left[\tilde{W}\gamma+\frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right]\left[\frac{\sigma_U(z)\tilde{\sigma}(z)}{\sigma^3(z)}\varepsilon+\frac{\rho_V\sigma_V\sigma_U(z)}{\sigma(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right]^2\right\}\times$$

$$\frac{2}{\tilde{\sigma}(z)}\Phi\left(-\lambda(z)\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right)\right)\phi\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right)d\varepsilon \qquad (I)$$

$$-\Phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\int_{\varepsilon}\frac{\Phi\left(-\frac{\tilde{W}\gamma}{\sqrt{1-\rho_V^2}}-\frac{\rho_V}{\sqrt{1-\rho_V^2}}\frac{\varepsilon}{\sigma_V}\right)}{\Phi(-\tilde{W}\gamma)}\frac{1}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)\times$$

$$\left[\frac{\rho_V^2\sigma_U(z)\left(\sigma_U^4(z)+3\sigma_U^2(z)\sigma_V^2-3\sigma_V^4\right)\varepsilon}{\sigma^6(z)}+\frac{\rho_V^2\sigma_V^4\sigma_U(z)}{\sigma^8(z)}\varepsilon^3\right]d\varepsilon \qquad (II)$$

$$-\phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\int_{\varepsilon}\frac{1}{\tilde{\sigma}_V}\phi\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}_V}\right)$$

$$\left\{-2\frac{\rho_V\sigma_V\tilde{\sigma}_V^2\sigma_U(z)}{\sigma^4(z)}\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)+\frac{\rho_V\sigma_V(1-\rho_V^2)\sigma_U(z)}{\tilde{\sigma}^2(z)}+\frac{\rho_V^3(1-\rho_V^2)\sigma_V^4\sigma_U^3(z)}{\tilde{\sigma}(z)\sigma^5(z)}\right.$$

$$\left.-\varepsilon^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\sigma^6(z)}-\tilde{W}\gamma\varepsilon\frac{(1-\rho_V^2)\sigma_V^2\sigma_U(z)}{\sigma^4(z)}-(\tilde{W}\gamma)^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\tilde{\sigma}^2(z)\sigma^2(z)}\right\}d\varepsilon \qquad (III)$$

We now treat each one each of these terms. Let

$$h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)=\Phi\left(-\lambda(z)\left(\frac{\varepsilon+\rho_V\sigma_V\tilde{W}\gamma}{\tilde{\sigma}(z)}\right)\right)\frac{2}{\tilde{\sigma}(z)}\phi\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}(z)}\right),$$

be the density of a $SN(-\tilde{W}\gamma\rho_V\sigma_V,\tilde{\sigma}(z),-\lambda(z))$. Therefore,

$$\int_{\varepsilon}\varepsilon h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon=-\tilde{W}\gamma\rho_V\sigma_V-\sqrt{\frac{2}{\pi}}\sigma_U(z),$$

$$\int_{\varepsilon}\varepsilon^2 h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon=\tilde{\sigma}^2(z)+(\tilde{W}\gamma)^2\rho_V^2\sigma_V^2+2\sqrt{\frac{2}{\pi}}\tilde{W}\gamma\rho_V\sigma_V\sigma_U(z),$$

$$\int_{\varepsilon}\varepsilon^3 h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon=-3\tilde{\sigma}^2(z)\tilde{W}\gamma\rho_V\sigma_V-3\sqrt{\frac{2}{\pi}}\tilde{\sigma}^2(z)\sigma_U(z)-(\tilde{W}\gamma)^3\rho_V^3\sigma_V^3$$

$$-3\sqrt{\frac{2}{\pi}}(\tilde{W}\gamma)^2\rho_V^2\sigma_V^2\sigma_U(z)+\sqrt{\frac{2}{\pi}}\sigma_U^3(z).$$

Thus, we have that

$$I=\phi(-\tilde{W}\gamma)\int_{\varepsilon}\left\{2\left[\frac{\rho_V\sigma_V\sigma_U^2(z)\tilde{\sigma}^2(z)}{\sigma^6(z)}\varepsilon+\frac{\rho_V^2\sigma_V^2\sigma_U^2(z)}{\sigma^4(z)}\tilde{W}\gamma\right]\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)\right.$$

$$+\frac{\rho_V\sigma_V\sigma_U^2(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}\left[3\rho_V^4\sigma_V^4-5\rho_V^2\sigma_V^2\sigma^2(z)+\sigma^4(z)\right]\varepsilon-\frac{\sigma_U^2(z)}{\tilde{\sigma}^2(z)\sigma^4(z)}\left[\tilde{\sigma}^4(z)+\rho_V^2\sigma_V^2\sigma_U^2(z)\right]\tilde{W}\gamma$$

$$\left.+\left[\tilde{W}\gamma+\frac{\rho_V\sigma_V}{\sigma^2(z)}\varepsilon\right]\left[\frac{\sigma_U(z)\tilde{\sigma}(z)}{\sigma^3(z)}\varepsilon+\frac{\rho_V\sigma_V\sigma_U(z)}{\sigma(z)\tilde{\sigma}(z)}\tilde{W}\gamma\right]^2\right\}h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon$$

$$=\phi(-\tilde{W}\gamma)\left\{-(\tilde{W}\gamma)^3\frac{\rho_V^6\sigma_V^6\sigma_U^2(z)}{\sigma^8(z)}+\sqrt{\frac{2}{\pi}}(\tilde{W}\gamma)^2\left[\frac{3\rho_V^7\sigma_V^7\sigma_U^3(z)}{\tilde{\sigma}^2(z)\sigma^8(z)}-\frac{4\rho_V^5\sigma_V^5\sigma_U^3(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}\right]\right.$$

$$+\tilde{W}\gamma\left[3\frac{\rho_V^5\sigma_V^5\sigma_U^2(z)}{\sigma^8(z)}-3\frac{\rho_V^2\sigma_V^2\sigma_U^2(z)\tilde{\sigma}^4(z)}{\sigma^8(z)}+\frac{6}{\pi}\frac{\rho_V^2\sigma_V^2\sigma_U^4(z)\tilde{\sigma}^2(z)}{\sigma^8(z)}\right]$$
$$+\sqrt{\frac{2}{\pi}}\left[-\frac{3\rho_V^7\sigma_V^7\sigma_U^3(z)}{\tilde{\sigma}^2(z)\sigma^8(z)}+\frac{4\rho_V^5\sigma_V^5\sigma_U^3(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}-\frac{\rho_V\sigma_V\sigma_U^5(z)\tilde{\sigma}^2(z)}{\sigma^8(z)}\right]\Bigg\}.$$

Similarly,

$$h_\varepsilon(\varepsilon)=\frac{\Phi\left(-\frac{\tilde{W}\gamma}{\sqrt{1-\rho_V^2}}-\frac{\rho_V}{\sqrt{1-\rho_V^2}}\frac{\varepsilon}{\sigma_V}\right)}{\Phi(-\tilde{W}\gamma)}\frac{1}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right),$$

is the density of an extended skew-normal distribution, $SN(0,\sigma_V,-\frac{\rho_V}{\sqrt{1-\rho_V^2}},-\tilde{W}\gamma)$, which implies that

$$\int_\varepsilon \varepsilon h_\varepsilon(\varepsilon)=-\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)}\rho_V\sigma_V$$
$$\int_\varepsilon \varepsilon^3 h_\varepsilon(\varepsilon)=-\frac{\phi(-\tilde{W}\gamma)}{\Phi(-\tilde{W}\gamma)}\left[\rho_V^3\sigma_V^3\left((\tilde{W}\gamma)^2-1\right)+3\rho_V\sigma_V^3\right)$$

Thus,

$$II=-\Phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\int_\varepsilon\frac{\Phi\left(-\frac{\tilde{W}\gamma}{\sqrt{1-\rho_V^2}}-\frac{\rho_V}{\sqrt{1-\rho_V^2}}\frac{\varepsilon}{\sigma_V}\right)}{\Phi(-\tilde{W}\gamma)}\frac{1}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)\times$$
$$\left[\frac{\rho_V^2\sigma_U(z)\left(\sigma_U^4(z)+3\sigma_U^2(z)\sigma_V^2-3\sigma_V^4\right)\varepsilon}{\sigma^6(z)}+\frac{\rho_V^2\sigma_V^4\sigma_U(z)}{\sigma^8(z)}\varepsilon^3\right]d\varepsilon$$
$$=\phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\left[(\tilde{W}\gamma)^2\frac{\rho_V^5\sigma_V^7\sigma_U(z)}{\sigma^8(z)}-\frac{\rho_V^5\sigma_V^7\sigma_U(z)}{\sigma^8(z)}+\frac{\rho_V^3\sigma_V\sigma_U^7(z)}{\sigma^8(z)}+\frac{4\rho_V^3\sigma_V^3\sigma_U^5(z)}{\sigma^8(z)}\right].$$

Finally, let

$$h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)=\frac{1}{\tilde{\sigma}_V}\phi\left(\frac{\varepsilon+\tilde{W}\gamma\rho_V\sigma_V}{\tilde{\sigma}_V}\right),$$

be the density of a $N(-\tilde{W}\gamma\rho_V\sigma_V,\tilde{\sigma}_V^2)$. Therefore,

$$\int_\varepsilon \varepsilon h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon=-\tilde{W}\gamma\rho_V\sigma_V$$
$$\int_\varepsilon \varepsilon^2 h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)d\varepsilon=\tilde{\sigma}_V^2+(\tilde{W}\gamma)^2\rho_V^2\sigma_V^2,$$

and

$$III=\phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\int_\varepsilon h_{\varepsilon|\tilde{W}\gamma}(\varepsilon|\tilde{W}\gamma)\left\{-2\frac{\rho_V\sigma_V\tilde{\sigma}_V^2\sigma_U(z)}{\sigma^4(z)}\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)+\frac{\rho_V\sigma_V(1-\rho_V^2)\sigma_U(z)}{\tilde{\sigma}^2(z)}\right.$$

$$+\frac{\rho_V^3(1-\rho_V^2)\sigma_V^4\sigma_U^3(z)}{\tilde{\sigma}(z)\sigma^5(z)}-\varepsilon^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\sigma^6(z)}$$

$$\left.-\tilde{W}\gamma\varepsilon\frac{(1-\rho_V^2)\sigma_V^2\sigma_U(z)}{\sigma^4(z)}-(\tilde{W}\gamma)^2\frac{\rho_V(1-\rho_V^2)\sigma_V^3\sigma_U(z)}{\tilde{\sigma}^2(z)\sigma^2(z)}\right\}d\varepsilon$$

$$=\phi(-\tilde{W}\gamma)\sqrt{\frac{2}{\pi}}\left\{-(\tilde{W}\gamma)^2\frac{\rho_V^5(1-\rho_V^2)\sigma_V^7\sigma_U(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}+\frac{\rho_V(1-\rho_V^2)\sigma_V\sigma_U^5(z)}{\tilde{\sigma}^2(z)\sigma^4(z)}\right.$$

$$\left.+\frac{\rho_V^5(1-\rho_V^2)\sigma_V^7\sigma_U(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}-\frac{\rho_V^3(1-\rho_V^2)\sigma_V^3\sigma_U^5(z)}{\tilde{\sigma}^2(z)\sigma^6(z)}\right\}$$

Therefore

$$\int_\varepsilon\left(\frac{\partial^2\Psi_0(z,\theta)}{\partial\rho_U^2}+\frac{\partial\Psi_0(z,\theta)}{\partial\rho_U}\frac{2\rho_V\sigma_V\sigma_U(z)}{\sigma^2(z)}\left(1-\frac{\varepsilon^2}{\sigma^2(z)}\right)\right)\frac{2}{\sigma(z)}\phi\left(\frac{\varepsilon}{\sigma(z)}\right)d\varepsilon$$

$$=\phi(-\tilde{W}\gamma)\left\{-(\tilde{W}\gamma)^3\frac{\rho_V^6\sigma_V^6\sigma_U^2(z)}{\sigma^8(z)}-(\tilde{W}\gamma)^2 4\sqrt{\frac{2}{\pi}}\frac{\rho_V^5\sigma_V^5\sigma_U^3(z)}{\sigma^8(z)}\right.$$

$$\left.+\tilde{W}\gamma\frac{3\rho_V^6\sigma_V^6\sigma_U^2(z)}{\sigma^8(z)}+4\sqrt{\frac{2}{\pi}}\frac{\rho_V^5\sigma_V^5\sigma_U^3(z)}{\sigma^8(z)}+4\sqrt{\frac{2}{\pi}}\frac{\rho_V^3\sigma_V^3\sigma_U^5(z)}{\sigma^8(z)}\right\}.$$

Now, we have

$$\left.\frac{\partial\Phi(\tau_1(z)\varepsilon)}{\partial\rho_U}\right|_{\rho_U=0}=\frac{\rho_V\sigma_V^2\varepsilon}{\sigma^3(z)}\phi\left(-\frac{\sigma_U(z)\varepsilon}{\sigma_V\sigma(z)}\right)$$

$$\left.\frac{\partial^2\Phi(\tau_1(z)\varepsilon)}{\partial\rho_U^2}\right|_{\rho_U=0}=\left(-\frac{\sigma_U(Z)\rho_V^2\left(\sigma_U^4(Z)+3\sigma_U^2(Z)\sigma_V^2-\sigma_V^4\right)\varepsilon}{\sigma_V\sigma^5(z)}+\frac{\rho_V^2\sigma_V^4\sigma_U(Z)\varepsilon^3}{\sigma_V\sigma^7(Z)}\right)\phi\left(-\frac{\sigma_U(z)\varepsilon}{\sigma_V\sigma(z)}\right),$$

so that

$$E_{\tilde{W}}\left[\int_\varepsilon\left(-\frac{\sigma_U(Z)\rho_V^2\left(\sigma_U^4(Z)+3\sigma_U^2(Z)\sigma_V^2-\sigma_V^4\right)\varepsilon}{\sigma_V\sigma^5(Z)}+\frac{\rho_V^2\sigma_V^4\sigma_U(Z)\varepsilon^3}{\sigma_V\sigma^7(Z)}\right.\right.$$

$$\left.\left.+\frac{2\rho_V^2\sigma_V^4\sigma_U(Z)\varepsilon}{\sigma_V\sigma^5(Z)}-\frac{2\rho_V^2\sigma_V^4\sigma_U(Z)\varepsilon^3}{\sigma_V\sigma^7(Z)}\right)\phi\left(-\frac{\sigma_U(Z)\varepsilon}{\sigma_V\sigma(Z)}\right)\frac{2}{\sigma(Z)}\phi\left(\frac{\varepsilon}{\sigma(Z)}\right)d\varepsilon\right]$$

$$=-\sqrt{\frac{2}{\pi}}E_{\tilde{W}}\left[\frac{\sigma_U(Z)\rho_V^2\left(\sigma_U^4(Z)+3\sigma_U^2(Z)\sigma_V^2-3\sigma_V^4\right)}{\sigma^6(Z)}\int_\varepsilon\frac{\varepsilon}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)d\varepsilon\right]$$

$$-\sqrt{\frac{2}{\pi}}E_{\tilde{W}}\left[\frac{\rho_V^2\sigma_V^4\sigma_U(Z)}{\sigma^8(Z)}\int_\varepsilon\frac{\varepsilon^3}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)d\varepsilon\right]=0,$$

where the final result follows because

$$\int_\varepsilon \frac{\varepsilon}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)d\varepsilon = 0,$$
$$\int_\varepsilon \frac{\varepsilon^3}{\sigma_V}\phi\left(\frac{\varepsilon}{\sigma_V}\right)d\varepsilon = 0,$$

by the properties of the pdf of a standard normal distribution. Thus, we can finally write

$$\begin{aligned}
SD_2 =& -\frac{\rho_V^6\sigma_V^6\sigma_U^2(1)}{\sigma^8(1)}E_{\tilde{W}}\left[\left(3\tilde{W}\gamma - (\tilde{W}\gamma)^3\right)\phi(-\tilde{W}\gamma)\right] \\
&+ 4\sqrt{\frac{2}{\pi}}\frac{\rho_V^5\sigma_V^5\sigma_U^3(1)}{\sigma^8(1)}E_{\tilde{W}}\left[\left((\tilde{W}\gamma)^2 - 1\right)\phi(-\tilde{W}\gamma)\right] \\
&- 4\sqrt{\frac{2}{\pi}}\frac{\rho_V^3\sigma_V^3\sigma_U^5(1)}{\sigma^8(1)}E_{\tilde{W}}\left[\phi(-\tilde{W}\gamma)\right] \\
SD_3 =& \frac{\rho_V^6\sigma_V^6\sigma_U^2(0)}{\sigma^8(0)}E_{\tilde{W}}\left[\left(3\tilde{W}\gamma - (\tilde{W}\gamma)^3\right)\phi(-\tilde{W}\gamma)\right] \\
&- 4\sqrt{\frac{2}{\pi}}\frac{\rho_V^5\sigma_V^5\sigma_U^3(0)}{\sigma^8(0)}E_{\tilde{W}}\left[\left((\tilde{W}\gamma)^2 - 1\right)\phi(-\tilde{W}\gamma)\right] \\
&+ 4\sqrt{\frac{2}{\pi}}\frac{\rho_V^3\sigma_V^3\sigma_U^5(0)}{\sigma^8(0)}E_{\tilde{W}}\left[\phi(-\tilde{W}\gamma)\right],
\end{aligned}$$

which implies that

$$\begin{aligned}
SD_2 + SD_3 =& -\left(\frac{\rho_V^6\sigma_V^6\sigma_U^2(1)}{\sigma^8(1)} - \frac{\rho_V^6\sigma_V^6\sigma_U^2(0)}{\sigma^8(0)}\right)E_{\tilde{W}}\left[\left(3\tilde{W}\gamma - (\tilde{W}\gamma)^3\right)\phi(-\tilde{W}\gamma)\right] \\
&+ 4\sqrt{\frac{2}{\pi}}\left(\frac{\rho_V^5\sigma_V^5\sigma_U^3(1)}{\sigma^8(1)} - \frac{\rho_V^5\sigma_V^5\sigma_U^3(0)}{\sigma^8(0)}\right)E_{\tilde{W}}\left[\left((\tilde{W}\gamma)^2 - 1\right)\phi(-\tilde{W}\gamma)\right] \\
&- 4\sqrt{\frac{2}{\pi}}\left(\frac{\rho_V^3\sigma_V^3\sigma_U^5(1)}{\sigma^8(1)} - \frac{\rho_V^3\sigma_V^3\sigma_U^5(0)}{\sigma^8(0)}\right)E_{\tilde{W}}\left[\phi(-\tilde{W}\gamma)\right],
\end{aligned}$$

and the final result follows from $SD_1 + SD_2 + SD_3 = 0$. This concludes the proof.

A.3. **Conditional density of U given ε.** From Centorrino and Pérez-Urdiales (2021), we have that

$$f_{U|\varepsilon,\eta}(u|\varepsilon,\eta) = \frac{1}{\sqrt{2\pi}\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\tilde{\sigma}(Z)}}\left\{\frac{f_{\varepsilon,\eta,1}(\varepsilon,\eta)}{f_{\varepsilon,\eta}(\varepsilon,\eta)}\left[\Phi\left(q_1(Z)\eta - \frac{\lambda(Z)}{\tilde{\sigma}(Z)}\varepsilon\right)\right]^{-1}\exp\left(-\frac{\left(u - \frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\tilde{\sigma}(Z)}\left(q_1(Z)\eta - \frac{\lambda(Z)}{\tilde{\sigma}(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde{\sigma}_V^2\tilde{\sigma}_U^2(Z)}{\tilde{\sigma}^2(Z)}}\right)\right.$$

$$+\frac{f_{\varepsilon,\eta,2}(\varepsilon,\eta)}{f_{\varepsilon,\eta}(\varepsilon,\eta)}\left[\Phi\left(q_2(Z)\eta-\frac{\lambda(Z)}{\tilde\sigma(Z)}\varepsilon\right)\right]^{-1}\exp\left(-\frac{\left(u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(q_2(Z)\eta-\frac{\lambda(Z)}{\tilde\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\tilde\sigma^2(Z)}}\right)\Bigg\}$$

$$=\frac{f_{\varepsilon,\eta,1}(\varepsilon,\eta)}{2f_{\varepsilon,\eta}(\varepsilon,\eta)}f_{U|\varepsilon,\eta,1}(u|\varepsilon,\eta)+\frac{f_{\varepsilon,\eta,2}(\varepsilon,\eta)}{2f_{\varepsilon,\eta}(\varepsilon,\eta)}f_{U|\varepsilon,\eta,2}(u|\varepsilon,\eta).$$

Let

$$f_{\varepsilon,1}(\varepsilon)=\frac{2}{\sqrt{2\pi(\tilde\sigma^2(Z)+\rho_1^2(Z))}}\Phi(\tau_1(Z)\varepsilon)\exp\left(-\frac{\varepsilon^2}{2(\tilde\sigma^2(Z)+\rho_1^2(Z))}\right)$$

$$f_{\varepsilon,2}(\varepsilon)=\frac{2}{\sqrt{2\pi(\tilde\sigma^2(Z)+\rho_2^2(Z))}}\Phi(\tau_2(Z)\varepsilon)\exp\left(-\frac{\varepsilon^2}{2(\tilde\sigma^2(Z)+\rho_2^2(Z))}\right),$$

which implies

$$f_\varepsilon(\varepsilon)=f_{\varepsilon,1}(\varepsilon)+f_{\varepsilon,2}(\varepsilon).$$

Using these notations, we can rewrite

$$f_{\eta|\varepsilon}(\eta|\varepsilon)=\frac{f_1(\varepsilon)}{f_\varepsilon(\varepsilon)}\frac{\Phi\left(q_1(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde\sigma(Z)}\right)}{\Phi(\tau_1(Z)\varepsilon)}\times$$
$$\sqrt{\frac{\tilde\sigma^2(Z)+\rho_1^2(Z)}{2\pi\tilde\sigma^2(Z)}}\exp\left(-\frac{\tilde\sigma^2(Z)+\rho_1^2(Z)}{2\tilde\sigma^2(Z)}\left(\eta-\frac{\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right)$$
$$+\frac{f_2(\varepsilon)}{f_\varepsilon(\varepsilon)}\frac{\Phi\left(q_2(Z)\eta-\frac{\lambda(Z)\varepsilon}{\tilde\sigma(Z)}\right)}{\Phi(\tau_2(Z)\varepsilon)}\times$$
$$\sqrt{\frac{\tilde\sigma^2(Z)+\rho_2^2(Z)}{2\pi\tilde\sigma^2(Z)}}\exp\left(-\frac{\tilde\sigma^2(Z)+\rho_2^2(Z)}{2\tilde\sigma^2(Z)}\left(\eta-\frac{\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right)$$
$$=\frac{f_1(\varepsilon)}{f_\varepsilon(\varepsilon)}f_{\eta|\varepsilon,1}(\eta|\varepsilon)+\frac{f_2(\varepsilon)}{f_\varepsilon(\varepsilon)}f_{\eta|\varepsilon,2}(\eta|\varepsilon).$$

We multiply the previous equation by the conditional density of U given (ε,η) to get

$$f_{U,\eta|\varepsilon}(u,\eta|\varepsilon)=f_{U|\varepsilon,\eta}(u|\varepsilon,\eta)f_{\eta|\varepsilon}(\eta|\varepsilon).$$

These computations give

$$f_{U,\eta|\varepsilon}(u,\eta|\varepsilon)=\frac{1}{2\pi\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}}\left\{\frac{f_1(\varepsilon)}{f_\varepsilon(\varepsilon)}\left[\Phi(\tau_1(Z)\varepsilon)\right]^{-1}\exp\left(-\frac{\left(u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(q_1(Z)\eta-\frac{\lambda(Z)}{\tilde\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\tilde\sigma^2(Z)}}\right)\times\right.$$

$$\sqrt{\frac{\tilde\sigma^2(Z)+\rho_1^2(Z)}{\tilde\sigma^2(Z)}}\exp\left(-\frac{\tilde\sigma^2(Z)+\rho_1^2(Z)}{2\tilde\sigma^2(Z)}\left(\eta-\frac{\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right)$$

$$+\frac{f_2(\varepsilon)}{f_\varepsilon(\varepsilon)}\left[\Phi\left(\tau_2(Z)\varepsilon\right)\right]^{-1}\exp\left(-\frac{\left(u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(q_2(Z)\eta-\frac{\lambda(Z)}{\tilde\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\tilde\sigma^2(Z)}}\right)$$

$$\left.\sqrt{\frac{\tilde\sigma^2(Z)+\rho_2^2(Z)}{\tilde\sigma^2(Z)}}\exp\left(-\frac{\tilde\sigma^2(Z)+\rho_2^2(Z)}{2\tilde\sigma^2(Z)}\left(\eta-\frac{\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right)\right\}.$$

Now, we let

$$\Sigma_{U\eta,1|\varepsilon}=\begin{bmatrix}\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\tilde\sigma^2(Z)}\left(1+\frac{q_1^2(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\right) & \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\frac{q_1(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\\ \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\frac{q_1(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)} & \frac{\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\end{bmatrix}$$

$$\Sigma_{U\eta,2|\varepsilon}=\begin{bmatrix}\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\tilde\sigma^2(Z)}\left(1+\frac{q_2^2(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\right) & \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\frac{q_2(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\\ \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\frac{q_2(Z)\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)} & \frac{\tilde\sigma^2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\end{bmatrix},$$

in a way that the expression above can be rewritten as

$$f_{U,\eta|\varepsilon}(u,\eta|\varepsilon)=$$

$$\frac{1}{2\pi\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}}\left\{\frac{f_1(\varepsilon)}{f_\varepsilon(\varepsilon)}\left[\Phi\left(\tau_1(Z)\varepsilon\right)\right]^{-1}\sqrt{\frac{\tilde\sigma^2(Z)+\rho_1^2(Z)}{\tilde\sigma^2(Z)}}\times\right.$$

$$\exp\left(-\frac{1}{2}\begin{pmatrix}u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(\frac{q_1(Z)\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}-\frac{\lambda(Z)}{\tilde\sigma(Z)}\right)\varepsilon\\ \eta-\frac{\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\end{pmatrix}'\Sigma_{U\eta,1|\varepsilon}^{-1}\begin{pmatrix}u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(\frac{q_1(Z)\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}-\frac{\lambda(Z)}{\tilde\sigma(Z)}\right)\varepsilon\\ \eta-\frac{\rho_1(Z)}{\tilde\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\end{pmatrix}\right)$$

$$+\frac{f_2(\varepsilon)}{f_\varepsilon(\varepsilon)}\left[\Phi\left(\tau_2(Z)\varepsilon\right)\right]^{-1}\sqrt{\frac{\tilde\sigma^2(Z)+\rho_2^2(Z)}{\tilde\sigma^2(Z)}}\times$$

$$\left.\exp\left(-\frac{1}{2}\begin{pmatrix}u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(\frac{q_2(Z)\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}-\frac{\lambda(Z)}{\tilde\sigma(Z)}\right)\varepsilon\\ \eta-\frac{\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\end{pmatrix}'\Sigma_{U\eta,2|\varepsilon}^{-1}\begin{pmatrix}u-\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\tilde\sigma(Z)}\left(\frac{q_2(Z)\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}-\frac{\lambda(Z)}{\tilde\sigma(Z)}\right)\varepsilon\\ \eta-\frac{\rho_2(Z)}{\tilde\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\end{pmatrix}\right)\right\}.$$

By integrating this joint conditional density with respect to η, we finally obtain

$$f_{U|\varepsilon}(u|\varepsilon) = \frac{f_{\varepsilon,1}(\varepsilon)}{f_{\varepsilon}(\varepsilon)} \frac{[\Phi(\tau_1(Z)\varepsilon)]^{-1}}{\sqrt{2\pi}\sigma_{1\star}} \exp\left(-\frac{(u-\mu_{1\star})^2}{2\sigma_{1\star}^2}\right) + \frac{f_{\varepsilon,2}(\varepsilon)}{f_{\varepsilon}(\varepsilon)} \frac{[\Phi(\tau_2(Z)\varepsilon)]^{-1}}{\sqrt{2\pi}\sigma_{2\star}} \exp\left(-\frac{(u-\mu_{2\star})^2}{2\sigma_{2\star}^2}\right),$$

which is a mixture of two half-normal densities, with weights given by

$$\omega_1(\varepsilon) = \frac{f_{\varepsilon,1}(\varepsilon)}{f_{\varepsilon}(\varepsilon)}, \qquad \omega_2(\varepsilon) = \frac{f_{\varepsilon,2}(\varepsilon)}{f_{\varepsilon}(\varepsilon)}.$$

Hence

$$E[\exp(-U)|\varepsilon] = \left\{\omega_1(\varepsilon)[\Phi(\tau_1(Z)\varepsilon)]^{-1} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_{1\star}} \exp\left(-u - \frac{(u-\mu_{1\star})^2}{2\sigma_{1\star}^2}\right) du + \omega_2(\varepsilon)[\Phi(\tau_2(Z)\varepsilon)]^{-1} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_{2\star}} \exp\left(-u - \frac{(u-\mu_{2\star})^2}{2\sigma_{2\star}^2}\right) du\right\}.$$

The final expression in (7) follows by the properties of the cdf of the univariate normal distribution.

Appendix B. Additional material for empirical application

In this section, we provide some additional information about the empirical application.

Table 5 contains descriptive statistics from the main variables used in the analysis. The variables are divided by category for convenience of the reader. Table 6 contains the MLE of the assignment equation with their 95% confidence intervals.

	Mean	St.Dev.	Min	Max
Output	1198.647	964.018	26.000	6816.000
Inputs				
Land	1.732	1.821	0.250	28.000
Labor	75.505	51.461	7.000	484.000
Fertilizers	209.037	159.201	9.000	1945.500
Pesticides	111.615	112.140	2.000	1116.600
Seeds	62.546	56.921	1.320	750.000
Environmental variables				
Tenure	0.658	0.475	0.000	1.000
Age	0.257	0.438	0.000	1.000
Education	7.022	3.391	0.000	16.000
No Income	0.673	0.470	0.000	1.000
Foot access	0.488	0.500	0.000	1.000
Car access	0.466	0.499	0.000	1.000
Risk div	1.199	1.758	-5.828	5.513
Participation	0.468	0.500	0.000	1.000
Instruments				
Dist Earthquake	0.756	0.430	0.000	1.000
Wage canton	4.060	0.987	0.000	6.000
Prop of families with electricity	0.820	0.232	0.000	1.000
Prop of families with bathroom	0.776	0.258	0.000	1.000

Table 5. Descriptive Statistics

	Estimate	CI	
γ_0	-2.3912	[-4.2259	-0.7052]
γ_{Land}	0.2533	[-0.0217	0.5675]
γ_{Labor}	-0.0076	[-0.1971	0.1926]
$\gamma_{Fertilizer}$	0.0474	[-0.1482	0.2324]
$\gamma_{Pesticides}$	0.0379	[-0.1399	0.2061]
γ_{Seeds}	0.1117	[-0.0428	0.2890]
γ_{Tenure}	0.2303	[-0.1764	0.6603]
$\gamma_{Tenure \times Land}$	-0.2096	[-0.5034	0.0411]
γ_{Age}	0.3106	[0.0203	0.6091]
γ_{Educ}	0.2021	[0.0768	0.3345]
$\gamma_{NoIncome}$	0.1490	[-0.1199	0.4328]
$\gamma_{FootAccess}$	0.2833	[-0.3062	0.8727]
$\gamma_{CarAccess}$	0.0490	[-0.5310	0.6455]
$\gamma_{RiskDiv}$	-0.0221	[-0.1467	0.1099]
$\gamma_{DistEarthquake}$	-0.8975	[-1.2014	-0.6175]
γ_{Wage}	0.0637	[-0.0659	0.2036]
$\gamma_{FamElect}$	0.7387	[0.1938	1.4231]
$\gamma_{FamBath}$	-0.4484	[-0.9240	0.0658]
γ_{Reg2}	0.7886	[0.4113	1.1983]
γ_{Reg3}	-0.1212	[-0.4182	0.1838]
γ_{Reg4}	1.7186	[1.3361	2.1848]

TABLE 6. *Estimation of the first-stage equation.*