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 efficiency of the agricultural sector is part of one of the Sustainable Development Goals set by the United Nations. Many international organizations have funded training programs that aim to reach small farmers in developing countries within this objective. Stochastic production frontier analysis can be a useful tool when evaluating the effectiveness of these programs. However, it does not allow one to take into account the selection bias often intrinsic to these interventions. In this work, we extend the classical maximum likelihood estimation of stochastic production frontier models, when both the inefficiency and the stochastic error term are affected by a potentially endogenous binary treatment variable. 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: August 24, 2020.
Authors wish to thank participants to vNAPW XI, especially Subal C. Kumbhakar, and to the Brown Bag Seminar at the Center of Productivity and Performance of Loughborough University for useful comments and remarks.

1. Introduction

The global need for improving efficiency in the agricultural sector has been recognized in the United Nation's 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. They generally imply either public investments, such as agricultural education and training or access to R&D, or price interventions, such as subsidies, affecting production prices (see, e.g. de Janvry et al., 2017, for a review).

However, participation in these programs often occurs voluntarily, which may lead to Selection Bias. 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 selection bias is not controlled for, one would conclude that the program is not effective because the agricultural efficiency of those who participate is lower than those who do not.

Moreover, Stochastic Frontier Analysis is a popular method to assess agricultural efficiency as the production potential of the agricultural sector. However, selection bias is particularly challenging to address within this framework.

The problem of selection bias in stochastic frontier model has been previously tackled in the literature from a methodological point of view. Greene (2010) formulates a selectivity corrected stochastic frontier model estimated using maximum simulated likelihood. The central assumption in this paper is that the unobservables in the selection equation are correlated with those in the production (or cost) equation, but uncorrelated with the inefficiency term. This approach is further extended by Bravo-Ureta et al. (2012) by using propensity score matching techniques to control for biases arising from observable variables and then controlling for selection bias stemming from unobservable characteristics following Greene (2010).

Kumbhakar et al. (2009) consider a framework similar to Greene (2010) but assuming that the selection mechanism operates through the efficiency term instead. In particular, this paper takes into account the endogeneity of technology choice (conventional or organic farming) by jointly estimating technology and technology choice using a single-step maximum likelihood method. However, to the best of our knowledge, there does not exist a maximum likelihood framework which allows one to control for potential correlation between program participation and both unobservable characteristics of participants and their stochastic inefficiency. Chen et al. (2020) study a general model with binary endogenous treatment and mediator, that are potentially correlated with the composite error term, but their approach boils down to a propensity score assumption.

In this paper, we contribute to this literature by providing a model that allows one to control for endogeneity coming from both sources, i.e., the correlation between program participation and the stochastic inefficiency and unobservable characteristics. Our empirical strategy is to employ instrumental variables to construct an auxiliary assignment mechanism for program participation. We then propose a maximum likelihood estimation framework in which we jointly model the density of the first stage error and the density of the composite error term common to stochastic production frontier. We derive the maximum likelihood function in closed form, which allows us to use standard estimation and inference procedure, making the model rather straightforward to estimate and interpret. Our model is similar to the one proposed by Chen et al. (2020) as we impose distributional assumptions on both the inefficiency term and the stochastic component. However, our approach is based on a one-step maximum likelihood estimator, which is straightforward to implement and allows one to obtain an estimator of technical efficiency, 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 apply the proposed method to smallholder farm household data from El Salvador. The data consist of a sample of participants in an Environmental program promoting soil conservation practices, and a control group of non-participant farmers. In this example, 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 does not only generate an upward shift on the frontier but also 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 contains a description of the sample, and outlines our empirical results. Finally, Section 4 concludes.

2. Binary Treatment in Stochastic Production Frontier

2.1. **Model.** In economic theory, the production frontier is defined as the quantity of output produced for a given input mix. In practice, producers can fail to produce exactly at the frontier and may fall below it. 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.

All these elements are usually combined in a stochastic frontier regression model of the type

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

where Y is the logarithm of output, $m(X, Z, \beta)$ is the logarithm of the production function, which depends on some unknown parameter β, some production inputs, X, and other *environmental factors*, Z, and $\varepsilon = V - U$ is an 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, their managerial decisions) 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. In our setting, we can interpret Z as the dummy for participation in the program fostering soil-conservation. This variable can therefore affect both the production frontier and the inefficiency of the producer. This is a binary treatment variable that takes value 1 if the producer participates in the program and 0 otherwise. 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 follows

$$m(X, Z, \beta) = m_0(X, \beta_0) + Zm_1(X, \beta_1),$$

in a way that the frontier *shifts* from m_0 to $m_0 + m_1$ as the treatment variable changes from 0 to 1. Therefore, our model in (1) becomes

$$Y = m_0(X, \beta_0) + Zm_1(X, \beta_1) + V - U.$$

For instance, in the prevalent case in which the logarithm of the production function is linear (i.e., Cobb-Douglas), this modeling strategy leads to include in the production frontier 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). Note that we propose a general framework in which the endogenous participation variable is included as both frontier shifter and efficiency determinant. Our estimation strategy also applies to the less general but more common case in which participation is only included as environmental factor affecting the inefficiency.

Maximum likelihood estimation is a popular approach to obtain estimators of the parameter (β, δ) in a stochastic frontier framework (Kumbhakar and Lovell, 2003). Although heavily parametrized, the likelihood specification allows one to obtain an estimator of the variance of the inefficiency term and thus understand how far from the frontier each producer is.

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 cannot be 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 normal-half-normal stochastic frontier model would lead to an inconsistent estimator of (β, δ).[1] Our goal is to construct a maximum likelihood estimator, which generalizes the normal-half-normal stochastic frontier model when the treatment is allowed to be endogenous.

In econometrics, instrumental variables are 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). In practice, it may be appealing to use a simple instrumental variable model, which ignores the binary nature of Z. This can be written as

$$Y = m(X, Z, \beta) + V - U_0 g(Z, \delta) \tag{2}$$

$$Z = \tilde{W}\gamma + \eta, \tag{3}$$

where $\tilde{W} = (W, X)$ and η is an error term, with $W \in \mathbb{R}^q$, and $q \geq 1$. In this construction, we are effectively splitting Z into two parts: $\tilde{W}\gamma$, which is independent of (V, U_0); and η, which instead captures the correlation between Z and (V, U_0) (Wooldridge, 2010). Z enters the second-stage equation nonlinearly, so the usual approach used in linear instrumental variable models to obtain the predicted values of Z from the first stage and use them in a second stage maximum likelihood estimation instead on Z would not lead to a consistent estimation of (β, δ) (Wooldridge, 2015; Amsler et al., 2016). An alternative approach consists in noticing that all the correlation between Z and (V, U_0) is captured by η. The latter can be considered an omitted variable in the second stage. The so-called *control function* approach consists in first obtaining an estimator of the first stage error and then plugging it in the second stage, effectively solving the omitted variable bias (Wooldridge, 2015).

While this approach might work, it also has some drawbacks. First of all, it is not obvious how an estimator of η should be *plugged* into the second stage. Shall we plug it in the production function, $m(X, Z, \beta)$? Or shall we plug it in the scaling function, $g(Z, \delta)$? Moreover, starting from the heuristic model in Equation (2), we cannot construct a maximum likelihood estimator, which generalizes the normal-half-normal model. Such an estimator would require us to specify the

[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 which is also robust to endogeneity in the inputs is possible, although we do not tackle it in this paper (see Centorrino and Pérez-Urdiales, 2020).

distribution of η, in a way that η is independent of $\tilde{W}$. However, it is known that the error term in the linear probability models is heteroskedastic by construction so that its variance depends on $\tilde{W}$.

If we acknowledge the binary nature of Z, we can instead use a Probit specification to model the treatment assignment (i.e., the first stage equation). In particular, 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.[2] The main assumptions of the Probit model are that $\eta \sim N(0,1)$ and that $\tilde{W}$ is independent of η. As above, we also impose that all the correlation between Z and (V, U_0) has to happen through η. This implies that once we control for η, the dependence between Z and (V, U_0) disappears (Newey et al., 1999; Imbens and Newey, 2009).

A maximum likelihood model requires to specify further the dependence between η and (V, U_0). More formally, this is done by modeling the conditional density of (V, U_0) given η. In parallel with the standard stochastic production frontier, we assume that, if there is any dependence between V and U_0, this 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. More specifically, we assume that $V|\eta \sim N(\rho_V \sigma_V \eta, \sigma_V^2)$ and that $U_0|\eta \sim FN(\rho_U \sigma_U \eta, \sigma_U^2)$, where σ_V and σ_U are the scale parameters of the conditional distributions of V and U_0 given η, respectively; ρ_V and ρ_U capture the dependence between (V, U_0) and η, respectively; and FN stands for a folded normal distribution, whose conditional pdf is given by the following expression

$$f_{U_0|\eta}(u|\eta) = \frac{1}{\sqrt{2\pi(1-\rho_U^2)\sigma_U^2}} \left\{ \exp\left(-\frac{(u - \rho_U \sigma_U \eta)^2}{2(1-\rho_U^2)\sigma_U^2} \right) + \exp\left(-\frac{(u + \rho_U \sigma_U \eta)^2}{2(1-\rho_U^2)\sigma_U^2} \right) \right\}. \tag{4}$$

Centorrino and Pérez-Urdiales (2020) 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). We can appreciate how the pdf in Equation (4) reduces to the half-normal distribution when $\rho_U = 0$. That is, when the treatment is assigned independently of the efficiency

[2] One can think about the first stage equation in terms of a latent variable $Z^* = \tilde{W}\gamma + \eta$ such that Z takes value 1 if $Z^* \geq 0$ and 0 otherwise.

of the producer. In the following, we refer to ρ_U as a correlation parameter, with a slight abuse of terminology.

Maximum likelihood estimation in stochastic frontier models is usually based on the density of the composite error term $\varepsilon = V - U$. In this endogenous case, 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 (2020), we know that the conditional density of ε given η is equal to

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

where

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

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

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

Let $\theta = (\beta', \delta', \gamma', \sigma_U^2, \sigma_V^2, \rho_V, \rho_U)'$ be the parameter 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, 2020). When Z is binary, this is not possible, as the first stage error term η cannot be directly estimated from the data, and thus we 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

$$\begin{aligned} P(Y \leq y, Z = 0|\tilde{W}) =& P(m(x, Z, \beta) + \varepsilon \leq y, Z = 0|\tilde{W}) \\ =& P(\varepsilon \leq y - m(X, Z, \beta), \eta \leq -\tilde{W}\gamma) = F_{\varepsilon,\eta}\left(y - m(X, Z, \beta), -\tilde{W}\gamma\right), \end{aligned} \tag{5}$$

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 (5) 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

$$\partial_1 F_{\varepsilon,\eta}\left(y - m(X, Z, \beta), -\tilde{W}\gamma\right) = f_{\varepsilon|\eta\leq-\tilde{W}\gamma}(y - m(X, Z, \beta)|\eta \leq -\tilde{W}\gamma)\left(1 - \Phi(\tilde{W}\gamma)\right),$$

where ∂_1 denotes the derivative with respect to the first argument of the function. Finally, we can write

$$\begin{aligned} f_{\varepsilon|\eta\leq-\tilde{W}\gamma}(y - m(X, Z, \beta)|\eta \leq -\tilde{W}\gamma) =& \int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon,\eta|\eta\leq-\tilde{W}\gamma}(y - m(X, Z, \beta), \eta|\eta \leq -\tilde{W}\gamma)d\eta \\ =& \int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon|\eta}(y - m(X, Z, \beta)|\eta) f_{\eta|\eta\leq-\tilde{W}\gamma}(\eta|\eta \leq -\tilde{W}\gamma)d\eta \\ =& \frac{1}{1 - \Phi(\tilde{W}\gamma)} \int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon|\eta}(y - m(X, Z, \beta)|\eta)\phi(\eta)d\eta, \end{aligned}$$

where the second line follows from the conditional independence of Z and ε given η, and the third line follows from the fact that, conditionally on $\eta \leq -\tilde{W}\gamma$, η follows a normal distribution which is truncated above by $-\tilde{W}\gamma$, and ϕ is the pdf of a standard normal distribution. We finally have that

$$\partial_1 F_{\varepsilon,\eta}\left(y - m(X,Z,\beta), -\tilde{W}\gamma\right) = \int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon|\eta}(y - m(X,Z,\beta)|\eta)\phi(\eta)d\eta.$$

The likelihood function can thus be obtained as

$$\mathcal{L}(\theta) = \left(\int_{-\tilde{W}\gamma}^{\infty} f_{\varepsilon|\eta}(y - m(X,Z,\beta)|\eta)\phi(\eta)d\eta\right)^{Z}\left(\int_{-\infty}^{-\tilde{W}\gamma} f_{\varepsilon|\eta}(y - m(X,Z,\beta)|\eta)\phi(\eta)d\eta\right)^{1-Z}, \qquad (6)$$

where the parameter $\theta = (\beta, \gamma, \delta, \rho_U, \rho_V, \sigma_U, \sigma_V)$.

The integrals appearing in the likelihood function can be solved analytically. A detailed derivation is provided in the Appendix. Here, we just summarize the main findings. We obtain that the conditional cdf of η is a mixture of two conditional skew normal distributions (Azzalini and Dalla Valle, 1996; Azzalini, 2013). When the correlation parameters ρ_V and ρ_U are equal to 0, this 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).

We thus have

$$\theta_0 = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta).$$

An important identification issue that arises in our framework is that only the magnitude of the parameter ρ_U is identified, but not its sign (see Centorrino and Pérez-Urdiales, 2020, for a detailed discussion). This implies that, while we can determine whether there is any dependence between the treatment assignment and inefficiency, we are not able to capture the sign of this dependence. As the parameter ρ_U is only identified up to a sign, the parameter θ_0 is not uniquely identified in the unrestricted parameter space Θ. However, we can still obtain local identification of θ_0 by properly restricting the parameter space (Sundberg, 1974). This lack of identification does not seem to be a crucial issue, and as shown in Centorrino and Pérez-Urdiales (2020), it does not have any major implications for estimation. It is however important when conducting inference in this setting, as discussed in their paper.

2.2. **Estimation.** 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

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} \left(\int_{-\tilde{W}_i\gamma}^{\infty} f_{\varepsilon|\eta}(\varepsilon_i|\eta)\phi(\eta)d\eta \right)^{Z_i} \left(\int_{-\infty}^{-\tilde{W}_i\gamma} f_{\varepsilon|\eta}(\varepsilon_i|\eta)\phi(\eta)d\eta \right)^{1-Z_i},$$

where $\varepsilon_i = Y_i - m(X_i, Z_i, \beta)$. Estimation is relatively straightforward, with the maximum likelihood estimator of the parameter θ given by

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

The likelihood function cannot be maximized analytically, and numerical optimization is required. This is usually achieved using Newton-Rapson iterations from a given starting value.

2.3. **Technical Efficiency.** A further step to complete our framework is to 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 (2020) 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. We let

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

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

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

where $\omega_l(\varepsilon)$, with $l = 1, 2$ are weights, such that $\omega_1(\varepsilon) + \omega_2(\varepsilon) = 1$. 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. Soil Conservation in El Salvador

3.1. **Data and Model Specification.** We consider data from an Environmental Program promoting crop diversification and soil conservation practices in El Salvador (PAES). The data set 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, in hillside plots with slopes of at least 15%. 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 detail on the program and data collection, see Bravo-Ureta et al. (2006).

For this application, we consider cross-sectional data focusing on the survey conducted in 2005. In a first instance, we estimate the following model

$$
Y_i = X_i\beta_1 + Z_{1i}\beta_2 + V_i - U_{0i}\exp(Z_i\delta), \text{ for } i = 1, \ldots, n, \tag{8}
$$

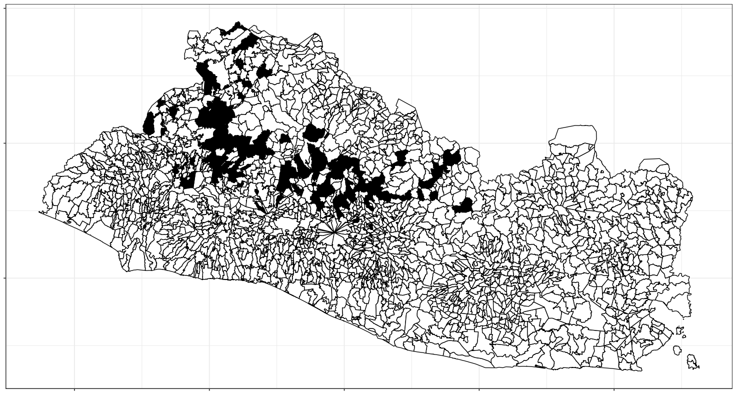

Figure 1. El Salvador - Location of cantons

where Y_i is the log of the total value of *Production* (measured in dollars); X_i is a vector of inputs including a constant term, the log of *Labour* (number of hired and household laborers), the log of *Land* (total area cultivated in manzana=0.7 Has), the log of *Fertilizers* (measured in dollars), the log of *Pesticides* (measured in dollars) and the log of *Seeds* (measured in dollars). $Z_i = (Z_{1i}, Z_{2i})$ is a random vector of environmental factors, which is decomposed into two sub-vectors: Z_{1i} can act both as a frontier shifter and as inefficiency determinant, while $Z_2 i$ only contains inefficiency determinants. Following Solis et al. (2009), among the frontier shifters, Z_{1i}, we consider the dummy variable *Slope* (dummy variable taking value 1 if the slope ≥15%, 0 otherwise) to control for differences in land quality, and *Participation* (dummy variable taking value 1 if the farmer is participating in PAES, 0 otherwise). Z_{2i}, instead, includes *Organization* (dummy variable equal to 1 if the farmer participates in a social organization, 0 otherwise), *Tenure* (percentage of land owned by the farmer over the total cultivated land), *Education* (measured in years), *Off-farm income* (measured in dollars), *Distance* (distance from the farmer's home to the closest plot, measured in km), *Risk div* (continuous index equal to 1 if the farmer cultivates only one crop, and therefore, not diversifying risks; or lower values if the farmer cultivates more crops).[3] *Slope* and *Participation* are also included as efficiency determinants, with the latter variable also interacted

[3]Following Centorrino and Pérez-Urdiales (2020), is constructed as follows:

$$Risk\ div = \begin{cases} \frac{\sum_{j=1}^{C} s_j^2 - 1/C}{1-1/C} & \text{for } C > 1 \\ 1 & \text{for } C = 1, \end{cases}$$

where s_j is the proportion of land devoted to crop j, and C is the total number of crops cultivated by each farmer. Our indicator ranges from 0 to 1.

with *Education*. V_i is the idiosyncratic error and U_{0i} is the stochastic inefficiency term. Our goal is to estimate the parameters $\{\alpha, \beta, \delta, \sigma_V^2, \sigma_U^2, \rho_U, \rho_V\}$. The total sample size is equal to $n = 443$.

In this application, *Participation* is an endogenous binary treatment variable. As instruments, we use *Distance_ city* (distance from the farmer's home to the closest city, measured in km), *Distance_ city*^{2} (square of the previous variable), *Electricity* (proportion of families in the farmer's canton with access to electricity), *Wage Canton* (the average hourly wage in the canton where the farmer is located), the interaction of the last two variables with *Distance_ city*, and the interaction between *Organization* and *Education* with *Distance_ city*. Descriptive statistics of the variables considered in the analysis are reported in Table 5 in Appendix.

We test the specification of the first stage regression (the regression of the treatment variable on all the other exogenous regressors) as a Probit model using the test proposed by 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 points out that our assignment mechanism is not misspecified. 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 to the instruments are set to zero. The null hypothesis of the test is that the instruments do not jointly have a significant effect on treatment assignment. The value of the likelihood ratio statistic is 31.33 which leads to a rejection of the null hypothesis with a p-value equal to 0.0003. The estimated coefficients of the first-stage Probit estimation are given in the Appendix.

3.2. **Results.** Table 1 reports results for our empirical example. The first pair of columns shows the estimation results assuming exogeneity. That is, when we ignore the potential selection issue associated with *Participation*. In this case, we find that the standard stochastic frontier model does not detect any inefficiency, and the exogenous model reduces to the standard linear regression with normal errors. Moreover, the Hessian matrix for the stochastic frontier model is singular (see Lee, 1993), estimates and standard errors for the estimator of the variance of the inefficiency term σ_U^2, and the parameter δ cannot be obtained. For the rest of the coefficients related to the production function and σ_V^2, one obtains standard errors using a Gaussian likelihood function.

The estimation results controlling for endogeneity are reported in the second pair of columns. We find that there is significant inefficiency when we control for the potential selection bias associated with *Participation*.

Recall that in this model, the parameters ρ_V and ρ_U capture the dependence between program participation and the unobserved components of the stochastic frontier model. We notice that both parameters are significantly different from 0 at standard significance levels. Thus, our model suggests a strong dependence between participation and unobservables. Moreover, the estimated value of ρ_V is negative, which suggests that the idiosyncratic components of production are negatively correlated with participation.

The estimated coefficients for *Labor* and *Land* are significant with the expected positive sign. However, the remaining inputs, *Fertilizers*, *Pesticides* and *Seeds* do not seem to have a significant effect. In terms of the frontier shifters, the coefficient for *Slope* is negative but not significant, indicating that there exist no differences in terms of land quality, but *Participation* is positive and significant, i.e., participating in the soil conservation program generates a significant upward shift in the frontier.

Regarding the efficiency determinants, the coefficient for *tenure* is positive and significant, which indicates that a higher proportion of owned land by the farmer has a negative impact on the efficiency (as it increases inefficiency). One possible explanation for this effect is that owning higher proportion of land implies lower flexibility to adapt to changing conditions. The variable *Risk div* also has a negative impact on the efficiency, since its estimated coefficient is positive and significant. That is, the greater the risk diversification(i.e., the more crops cultivated by the farmer), the lower the level of efficiency, likely related to the lower levels of specialization and economies of scale that can be achieved. Surprisingly, the estimated coefficient for *Organization* is positive and significant, which implies that farmers belonging to an organization show lower levels of efficiency. The coefficient of our variable of interest in this study, *Participation*, is negative and significant, while the one of the interaction between this variable and *Education* is positive. This indicates that while participating in the soil conservation program improves efficiency (reduces inefficiency), the effect of program participation on efficiency decreases as the farmer's education increases. One possible explanation for this result is that more educated farmers may have already implemented some of the soil conservation measures proposed by the program before participating in it.

Using standard inference based on asymptotic normality, one may conclude that the estimator of the variance of the inefficiency in the endogenous model is not significantly different from 0. However, as the sample size is not large, the asymptotic approximation may not be very precise. To overcome this issue, we compute the 95% profile likelihood confidence interval for σ_U^2 (Cox and Hinkley, 1979). We find that the profile-likelihood confidence interval for the variance of the inefficiency term σ_U^2 (multiplied by 1000 for ease of reading) $CI(\sigma_U^2) = [0.0141, 0.0496]$, which shows that the confidence interval is much tighter than the standard errors computed from the inverse numerical Hessian would suggest. As the likelihood function cannot be maximized analytically, and numerical evaluation of the first and second derivative is necessary, a larger sample size may be required for obtaining smaller standard errors. A similar result holds for other parameters in the model, although we do not report their confidence intervals. This suggests that, in this context, profile-likelihood confidence intervals may provide a better tool for determining parameter uncertainty. This observation is in line with the small sample findings of Centorrino and Pérez-Urdiales (2020), who have shown that the likelihood ratio test performs better than the Wald test in this setting.

Last, technical efficiency is estimated for each farmer using Equation (7). The mean technical efficiency is reported at the bottom in Table 1, and Figure 2 provides the kernel density estimators of this measure for both the stochastic frontier model assuming exogeneity (dashed line) and the stochastic frontier model controlling for endogeneity (black line). The mean technical efficiency for the model controlling for endogeneity is equal to 0.8697. However, estimates of technical efficiency vary considerably across farmers, with the majority of farmers being highly efficient, and very few farmers being relatively inefficient.

4. Conclusions

In the current need for increasing agricultural efficiency, Stochastic Production Frontier Models can shed light on the effectiveness of programs targeting this goal. However, controlling for potential endogeneity associated to voluntary program participation is crucial to obtain an accurate estimate

	Exogeneity		Endogeneity	
	Estimate	Std. Err.	Estimate	Std. Err.
β_0	3.8953	0.1732	6.7913	0.5221
β_{Land}	0.3872	0.0397	0.3501	0.0501
β_{Labour}	0.6071	0.0417	0.5652	0.0539
$\beta_{Fertilizer}$	0.0224	0.0216	0.0247	0.0255
$\beta_{Pesticides}$	-0.0111	0.0185	-0.0171	0.0217
β_{Seeds}	0.0574	0.0197	0.0386	0.0239
β_{Slope}	0.0779	0.0424	-0.0344	0.0540
$\beta_{Participation}$	-0.0654	0.0421	0.4937	0.0703
$\delta_{Organization}$			2.9861	1.2126
δ_{Educ}			-0.2322	0.3914
δ_{Tenure}			1.5870	0.9701
$\delta_{Off-farmincome}$			-0.2032	0.1527
$\delta_{Distance}$			-0.3920	0.3010
δ_{Slope}			-0.1999	0.5445
δ_{Div}			2.6483	1.1016
$\delta_{Educ \times Participation}$			2.0022	1.0438
$\delta_{Participation}$			-6.3459	3.1446
$\rho_{U,\eta}$			-0.8696	0.0266
$\rho_{V,\eta}$			-0.8297	0.1542
σ_U^2	0		0.0000	0.0001
σ_V^2	0.1813	0.0116	0.2582	0.0266
Mean TE	0.9952		0.8697	

TABLE 1. *Estimation of the efficiency frontier with and without accounting for treatment endogeneity.*

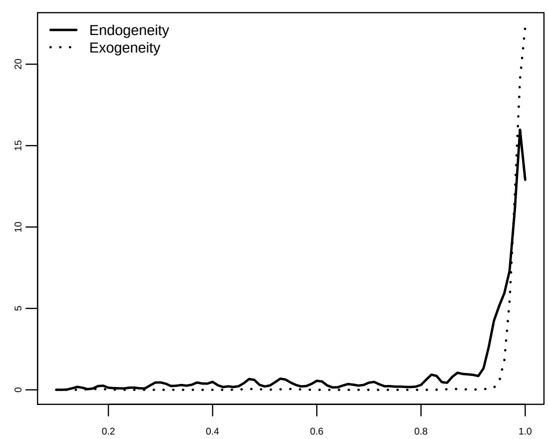

FIGURE 2. Estimation of technical efficiency.

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 construct a simple closed-form 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 models assuming exogeneity of program participation either do not detect inefficiency or do not find a significant effect of the program on the level of production and efficiency. However, when we implement the method proposed we find substantial production potential (as we detect a significant level of technical inefficiency) and that participation in the soil conservation program generated an upward shift on the frontier and a reduction in the level of technical inefficiency.

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. This is especially important in guiding future evidence-based policy-making in which the best possible use should be made of scarce resources.

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.

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.

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., 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. (2020), 'Maximum Likelihood Estimation of Stochastic Frontier Models with Endogeneity', *Mimeo - Stony Brook University* .

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.

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.
URL: *http://www.sciencedirect.com/science/article/pii/S2214658X16300046*

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

Fletcher, R. (1987), *Practical methods of optimization*, Wiley-Interscience publication, Wiley.

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

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

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., 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.

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

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.

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

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

Solis, D., Bravo-Ureta, B. E. and Quiroga, R. E. (2009), 'Technical efficiency among peasant farmers participating in natural resource management programmes in central america', *Journal of Agricultural Economics* **60**(1), 202–219.

URL: *https://doi.org/10.1111/j.1477-9552.2008.00173.x*

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

Vo Hung Son, T., Coelli, T. and Fleming, E. (1993), 'Analysis of the technical efficiency of state rubber farms in vietnam', *Agricultural Economics* **9**(3), 183 – 201.

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

Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, second edn, Boston MA: MIT Press.

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

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

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

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

(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

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

Therefore, we have that

$$
\begin{aligned}
f_{\varepsilon,\eta}(\varepsilon,\eta) = & \frac{1}{2\pi\sigma(Z)}\left\{\Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\sigma(Z)} + \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\sigma(Z)} - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)\times\right. \\
& \exp\left(-\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2 - \frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_1^2(Z))}\right) \\
& + \Phi\left(\frac{\lambda(Z)\rho_V\sigma_V\eta}{\sigma(Z)} - \frac{\rho_U\sigma_U(Z)\eta}{\lambda(Z)\sigma(Z)} - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)\times \\
& \left. \exp\left(-\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2 - \frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_2^2(Z))}\right)\right\}.
\end{aligned}
$$

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}{\sigma(Z)}\right) \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right) \times \\
& \frac{1}{\sqrt{2\pi(\sigma^2(Z)+\rho_1^2(Z))}} \exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_1^2(Z))}\right) \\
& + \Phi\left(q_2(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right) \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right) \times \\
& \frac{1}{\sqrt{2\pi(\sigma^2(Z)+\rho_2^2(Z))}} \exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_2^2(Z))}\right)\Bigg\} \\
= & \frac{1}{2}\Bigg\{\frac{\Phi\left(q_1(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)}{\Phi\left(\tau_1(Z)\varepsilon\right)} \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right) \times \\
& \frac{2}{\sqrt{2\pi(\sigma^2(Z)+\rho_1^2(Z))}} \Phi\left(\tau_1(Z)\varepsilon\right) \exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_1^2(Z))}\right) \\
& + \frac{\Phi\left(q_2(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)}{\Phi\left(\tau_2(Z)\varepsilon\right)} \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right) \times \\
& \frac{2}{\sqrt{2\pi(\sigma^2(Z)+\rho_2^2(Z))}} \Phi\left(\tau_2(Z)\varepsilon\right) \exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z)+\rho_2^2(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

$$q_1(Z) = \frac{\lambda(Z)\rho_V\sigma_V}{\sigma(Z)} + \frac{\rho_U\sigma_U(Z)}{\lambda(Z)\sigma(Z)}$$

$$q_2(Z) = \frac{\lambda(Z)\rho_V\sigma_V}{\sigma(Z)} - \frac{\rho_U\sigma_U(Z)}{\lambda(Z)\sigma(Z)}$$

$$\tau_1(Z) = \frac{\frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)} q_1(Z) - \frac{\lambda(Z)}{\sigma(Z)}}{\sqrt{1 + \frac{q_1^2(Z)\sigma^2(Z)}{\sigma^2(Z)+\rho_1^2(Z)}}}$$

$$\tau_2(Z) = \frac{\frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)} q_2(Z) - \frac{\lambda(Z)}{\sigma(Z)}}{\sqrt{1 + \frac{q_2^2(Z)\sigma^2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}}},$$

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

$$\mathcal{I}_j = \int_{a_j}^{b_j} \Phi\left(q_j(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right) \times \sqrt{\frac{\sigma^2(Z)+\rho_j^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_j^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_j(Z)}{\sigma^2(Z)+\rho_j^2(Z)}\varepsilon\right)^2\right) d\eta,$$

for $j = 1, 2$, cannot be computed in closed form. However, using the properties of the skew normal distribution, it can be expressed as the cdf of a bivariate normal distribution.

With a slight abuse of notations, let us define the fictitious random variables $(\eta_j, \varepsilon_j, \kappa_j)$, for $j = 1, 2$, such that the conditional distribution of (κ_j, η_j) given ε_j is a bivariate normal distribution with mean and variance given by

$$\mu_j(Z) = \begin{pmatrix} \tau_j(Z)\varepsilon_j \\ \frac{\rho_j(Z)}{\sigma^2(Z)+\rho_j^2(Z)}\varepsilon_j \end{pmatrix}, \qquad \Omega_j(Z) = \begin{pmatrix} 1 & -\frac{\frac{q_j(\xi)\sigma^2(Z)}{\sigma^2(Z)+\rho_j^2(Z)}}{\sqrt{1+\frac{q_j^2(Z)\sigma^2(Z)}{\sigma^2(Z)+\rho_l^2(Z)}}} \\ -\frac{\frac{q_j(\xi)\sigma^2(Z)}{\sigma^2(Z)+\rho_j^2(Z)}}{\sqrt{1+\frac{q_j^2(Z)\sigma^2(Z)}{\sigma^2(Z)+\rho_j^2(Z)}}} & \frac{\sigma^2(Z)}{\sigma^2(Z)+\rho_j^2(Z)} \end{pmatrix}.$$

Therefore, we have that

$$\mathcal{I}_j = \Phi_2\left(\begin{pmatrix} 0 \\ b_j \end{pmatrix} - \mu_j(Z), \Omega_j(Z)\right) - \Phi_2\left(\begin{pmatrix} 0 \\ a_j \end{pmatrix} - \mu_j(Z), \Omega_j(Z)\right),$$

for $j = 1,2$, where Φ_2 is the cdf of a bivariate normal distribution, with mean 0 and covariance matrix Ω. Ultimately, these integrals only involve numerical integration of a bivariate normal distribution, which is readily available in any standard statistical software.

To conclude, we notice that, when there is no endogeneity issue in this model, the likelihood function written above collapses to the product of two independent densities, so that the model as written here provides a natural generalization of standard stochastic frontier analysis (Centorrino and Pérez-Urdiales, 2020).

A.2. **Algorithm and properties of the maximum likelihood estimator.** Let $\hat{\theta}_n^{(0)} = (\hat{\beta}_n^{(0)}, \hat{\gamma}_n^{(0)}, \hat{\delta}_n^{(0)}, \hat{\rho}_{U,n}^{(0}, \hat{\rho}_{V,n}^{(0)}, \hat{\sigma}_{U,n}^{(0)}, \hat{\sigma}_{V,n}^{(0)})$ be a starting value for the parameter θ. We let

$$\hat{\sigma}_{U,n}^{(0)}(Z) = \hat{\sigma}_{U,n}^{(0)} g\left(Z, \hat{\delta}_n^{(0)}\right).$$

For $Z = \{0,1\}$, we let

$$\begin{aligned}
\hat{\rho}_1^{(0)}(Z_i) =& \hat{\rho}_{V,n}^{(0)}\hat{\sigma}_{V,n}^{(0)} - \hat{\rho}_{U,n}^{(0)}\hat{\sigma}_{U,n}^{(0)}(Z) \\
\hat{\rho}_2^{(0)}(Z_i) =& \hat{\rho}_{V,n}^{(0)}\hat{\sigma}_{V,n}^{(0)} + \hat{\rho}_{U,n}^{(0)}\hat{\sigma}_{U,n}^{(0)}(Z_i) \\
(\hat{\sigma}^{(0)}(Z_i))^2 =& \left(1 - (\hat{\rho}_{V,n}^{(0)})^2\right)(\hat{\sigma}_{V,n}^{(0)})^2 + \left(1 - (\hat{\rho}_{U,n}^{(0)})^2\right)(\hat{\sigma}_{U,n}^{(0)}(Z_i))^2 \\
\hat{\lambda}^{(0)}(Z_i) =& \frac{\sqrt{1 - (\hat{\rho}_{U,n}^{(0)})^2}\hat{\sigma}_{U,n}^{(0)}(Z_i)}{\sqrt{1 - (\hat{\rho}_{V,n}^{(0)})^2}\hat{\sigma}_{V,n}^{(0)}} \\
\hat{q}_1^{(0)}(Z_i) =& \frac{\hat{\lambda}^{(0)}(Z_i)\hat{\rho}_{V,n}^{(0)}\hat{\sigma}_{V,n}^{(0)}}{\hat{\sigma}^{(0)}(Z_i)} + \frac{\hat{\rho}_{U,n}^{(0)}\hat{\sigma}_{U,n}^{(0)}(Z_i)}{\hat{\lambda}^{(0)}(Z_i)\hat{\sigma}^{(0)}(Z_i)} \\
\hat{q}_2^{(0)}(Z_i) =& \frac{\hat{\lambda}^{(0)}(Z_i)\hat{\rho}_{V,n}^{(0)}\hat{\sigma}_{V,n}^{(0)}}{\hat{\sigma}^{(0)}(Z_i)} - \frac{\hat{\rho}_{U,n}^{(0)}\hat{\sigma}_{U,n}^{(0)}(Z_i)}{\hat{\lambda}^{(0)}(Z_i)\hat{\sigma}^{(0)}(Z_i)} \\
\hat{\tau}_1^{(0)}(Z_i) =& \sqrt{1 + \left(\frac{\hat{\lambda}^{(0)}(Z_i)\hat{\rho}_1^{(0)}(Z_i)}{\hat{\sigma}^{(0)}(Z_i)} - \hat{q}_1^{(0)}(Z_i)\right)^2 + (\hat{\lambda}^{(0)}(Z_i))^2} \\
\hat{\tau}_2^{(0)}(Z_i) =& \sqrt{1 + \left(\frac{\hat{\lambda}^{(0)}(Z_i)\hat{\rho}_2^{(0)}(Z_i)}{\hat{\sigma}^{(0)}(Z_i)} - \hat{q}_2^{(0)}(Z_i)\right)^2 + (\hat{\lambda}^{(0)}(Z_i))^2}.
\end{aligned}$$

This is used to construct the residual

$$\hat{\varepsilon}_i^{(0)} = Y_i - m(X_i, Z_i, \hat{\beta}_n^{(0)}).$$

These are then plugged into the likelihood function along with $\hat{\theta}_n^{(0)}$, to obtain the value of the function at the point $\hat{\theta}_n^{(0)}$ and its numerical first and second derivatives. The direction of the search for a maximum is then chosen using a Quasi-Newton algorithm (such as the Broyden–Fletcher–Goldfarb–Shanno, BFGS, algorithm, see e.g. Fletcher, 1987, among others). The search continues until the distance between consecutive iterations is smaller than a fixed tolerance level. In order to be robust with respect to the choice of the initial condition, we restart the optimization process from several different points.

If the true value of the parameter, θ_0 is in the interior of a proper partition of the parameter space, and the likelihood function is at least twice differentiable in a neighborhood of the true value, we have that

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

where $I\left(\theta_0\right)$ is the Fisher's information matrix at the true value of θ.

A.3. **Conditional density of U given ε.** From Centorrino and Pérez-Urdiales (2020), we have that

$$\begin{aligned} f_{U|\varepsilon,\eta}(u|\varepsilon,\eta) =& \frac{1}{2\sqrt{2\pi}\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}}\left\{\frac{f_{\varepsilon,\eta,1}\left(\varepsilon,\eta\right)}{f_{\varepsilon,\eta}\left(\varepsilon,\eta\right)}\left[\Phi\left(q_1(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right]^{-1}\exp\left(-\frac{\left(u - \frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(q_1(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde{\sigma}_V^2\tilde{\sigma}_U^2(Z)}{\sigma^2(Z)}}\right)\right. \\ &\left. + \frac{f_{\varepsilon,\eta,2}\left(\varepsilon,\eta\right)}{f_{\varepsilon,\eta}\left(\varepsilon,\eta\right)}\left[\Phi\left(q_2(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right]^{-1}\exp\left(-\frac{\left(u - \frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(q_2(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde{\sigma}_V^2\tilde{\sigma}_U^2(Z)}{\sigma^2(Z)}}\right)\right\} \\ =& \frac{f_{\varepsilon,\eta,1}\left(\varepsilon,\eta\right)}{2f_{\varepsilon,\eta}\left(\varepsilon,\eta\right)}f_{U|\varepsilon,\eta,1}(u|\varepsilon,\eta) + \frac{f_{\varepsilon,\eta,2}\left(\varepsilon,\eta\right)}{2f_{\varepsilon,\eta}\left(\varepsilon,\eta\right)}f_{U|\varepsilon,\eta,2}(u|\varepsilon,\eta). \end{aligned}$$

Let

$$\begin{aligned} f_{\varepsilon,1}\left(\varepsilon\right) =& \frac{2}{\sqrt{2\pi(\sigma^2(Z) + \rho_1^2(Z))}}\Phi\left(\tau_1(Z)\varepsilon\right)\exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z) + \rho_1^2(Z))}\right) \\ f_{\varepsilon,2}\left(\varepsilon\right) =& \frac{2}{\sqrt{2\pi(\sigma^2(Z) + \rho_2^2(Z))}}\Phi\left(\tau_2(Z)\varepsilon\right)\exp\left(-\frac{\varepsilon^2}{2(\sigma^2(Z) + \rho_2^2(Z))}\right), \end{aligned}$$

which implies

$$f_\varepsilon(\varepsilon) = 0.5\left(f_{\varepsilon,1}(\varepsilon) + f_{\varepsilon,2}(\varepsilon)\right).$$

Using these notations, we can rewrite

$$\begin{aligned}
f_{\eta|\varepsilon}(\eta|\varepsilon) =& \frac{f_1(\varepsilon)}{2f_\varepsilon(\varepsilon)} \frac{\Phi\left(q_1(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)}{\Phi\left(\tau_1(Z)\varepsilon\right)} \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right) \\
& + \frac{f_2(\varepsilon)}{2f_\varepsilon(\varepsilon)} \frac{\Phi\left(q_2(Z)\eta - \frac{\lambda(Z)\varepsilon}{\sigma(Z)}\right)}{\Phi\left(\tau_2(Z)\varepsilon\right)} \times \\
& \sqrt{\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\pi\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right) \\
=& \frac{f_1(\varepsilon)}{f_\varepsilon(\varepsilon)} f_{\eta|\varepsilon,1}(\eta|\varepsilon) + \frac{f_2(\varepsilon)}{f_\varepsilon(\varepsilon)} f_{\eta|\varepsilon,2}(\eta|\varepsilon)
\end{aligned} \tag{9}$$

We multiply Equation (9) 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

$$\begin{aligned}
f_{U,\eta|\varepsilon}(u, \eta|\varepsilon) =& \frac{1}{2\pi\frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\sigma(Z)}} \left\{ \frac{f_1(\varepsilon)}{2f_\varepsilon(\varepsilon)} \left[\Phi\left(\tau_1(Z)\varepsilon\right)\right]^{-1} \exp\left(-\frac{\left(u - \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\sigma(Z)}\left(q_1(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\sigma^2(Z)}}\right) \times \right. \\
& \sqrt{\frac{\sigma^2(Z)+\rho_1^2(Z)}{\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_1^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon\right)^2\right) \\
& + \frac{f_2(\varepsilon)}{2f_\varepsilon(\varepsilon)} \left[\Phi\left(\tau_2(Z)\varepsilon\right)\right]^{-1} \exp\left(-\frac{\left(u - \frac{\tilde\sigma_V\tilde\sigma_U(Z)}{\sigma(Z)}\left(q_2(Z)\eta - \frac{\lambda(Z)}{\sigma(Z)}\varepsilon\right)\right)^2}{2\frac{\tilde\sigma_V^2\tilde\sigma_U^2(Z)}{\sigma^2(Z)}}\right) \\
& \left. \sqrt{\frac{\sigma^2(Z)+\rho_2^2(Z)}{\sigma^2(Z)}} \exp\left(-\frac{\sigma^2(Z)+\rho_2^2(Z)}{2\sigma^2(Z)}\left(\eta - \frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon\right)^2\right) \right\}.
\end{aligned}$$

Now, we let

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

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

in a way that the expression above can be rewritten as

$$\begin{aligned}
&f_{U,\eta|\varepsilon}(u,\eta|\varepsilon) = \\
&\quad \frac{1}{2\pi\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}}\left\{\frac{f_1(\varepsilon)}{2f_\varepsilon(\varepsilon)}\left[\Phi\left(\tau_1(Z)\varepsilon\right)\right]^{-1}\sqrt{\frac{\sigma^2(Z)+\rho_1^2(Z)}{\sigma^2(Z)}}\times\right. \\
&\qquad \exp\left(-\frac{1}{2}\begin{pmatrix} u-\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(\frac{q_1(Z)\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}-\frac{\lambda(Z)}{\sigma(Z)}\right)\varepsilon \\ \eta-\frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon \end{pmatrix}'\Sigma_{U\eta,1|\varepsilon}^{-1}\begin{pmatrix} u-\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(\frac{q_1(Z)\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}-\frac{\lambda(Z)}{\sigma(Z)}\right)\varepsilon \\ \eta-\frac{\rho_1(Z)}{\sigma^2(Z)+\rho_1^2(Z)}\varepsilon \end{pmatrix}\right) \\
&\qquad +\frac{f_2(\varepsilon)}{2f_\varepsilon(\varepsilon)}\left[\Phi\left(\tau_2(Z)\varepsilon\right)\right]^{-1}\sqrt{\frac{\sigma^2(Z)+\rho_2^2(Z)}{\sigma^2(Z)}}\times \\
&\qquad \left.\exp\left(-\frac{1}{2}\begin{pmatrix} u-\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(\frac{q_2(Z)\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}-\frac{\lambda(Z)}{\sigma(Z)}\right)\varepsilon \\ \eta-\frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon \end{pmatrix}'\Sigma_{U\eta,2|\varepsilon}^{-1}\begin{pmatrix} u-\frac{\tilde{\sigma}_V\tilde{\sigma}_U(Z)}{\sigma(Z)}\left(\frac{q_2(Z)\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}-\frac{\lambda(Z)}{\sigma(Z)}\right)\varepsilon \\ \eta-\frac{\rho_2(Z)}{\sigma^2(Z)+\rho_2^2(Z)}\varepsilon \end{pmatrix}\right)\right\}.
\end{aligned}$$

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

$$f_{U|\varepsilon}(u|\varepsilon) = \left\{\frac{f_{\varepsilon,1}(\varepsilon)}{2f_\varepsilon(\varepsilon)}\frac{\left[\Phi\left(\tau_1(Z)\varepsilon\right)\right]^{-1}}{\sqrt{2\pi}\sigma_{1\star}}\exp\left(-\frac{(u-\mu_{1\star})^2}{2\sigma_{1\star}^2}\right)+\frac{f_{\varepsilon,2}(\varepsilon)}{2f_\varepsilon(\varepsilon)}\frac{\left[\Phi\left(\tau_2(Z)\varepsilon\right)\right]^{-1}}{\sqrt{2\pi}\sigma_{2\star}}\exp\left(-\frac{(u-\mu_{2\star})^2}{2\sigma_{2\star}^2}\right)\right\}, \tag{10}$$

which is a mixture of two half-normal densities, with weights given by $f_{\varepsilon,1}(\varepsilon)/(2f_\varepsilon(\varepsilon))$ and $f_{\varepsilon,2}(\varepsilon)/(2f_\varepsilon(\varepsilon))$.

Hence

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

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

B. Simulations

We replicate a similar simulation scheme as in Amsler et al. (2017) and Centorrino and Pérez-Urdiales (2020). 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 $\beta_1 = \beta_2 = 0.66074$, $\delta_1 = 0.05$ and $\delta_2 = -0.2$ 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 one continuous instrument W_i, also fully independent of the error term, and such that

$$Z_{2i} = \mathbb{1}\left(\gamma\left(X_{1i} + X_{2i} + Z_{1i} + W_i\right) + \eta \geq 0\right),$$

where $\mathbb{1}(\cdot)$ is the indicator function, and $\gamma = 0.31623$.

The exogenous variables are generated independently from a normal distribution with means equal to 0 and variances equal to 1. These variables are equicorrelated, with correlation parameter equal to 0.5.

We generate the pair (V, η) from the following 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 follows

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

where ϵ is a standard normal random variable.

We consider two simulation schemes that differ because of the value of the parameter ρ_U. In setting 1, we take U_0 to be uncorrelated with η. In setting 2, we take $\rho_U = 0.5$. We take increasing sample sizes $n = \{250, 500, 1000\}$, and run $R = 1000$ replications for each scenario.

Our estimation procedure is based on the maximization of the full likelihood in Equation (6). There are two main issues for practical implementation. First, the parameter space is often very large. In simulations, we maximize the log-likelihood function with respect to the full vector of parameters. To reduce the dimensionality of the optimization problem, one can estimate the vector of parameters γ by OLS. For a given γ, one can then maximize the full likelihood with respect to the other parameters. One can use the estimator obtained in this fashion as a starting value for maximization of the full likelihood. Standard errors are obtained by evaluating numerically the Hessian matrix of the full likelihood.

We report results of these simulations in Tables 2 and 3 below.

		N = 250		N = 500		N = 1000	
	TRUE	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
β_0	0.0000	-0.0985	0.3077	-0.0492	0.1993	-0.0182	0.1169
β_1	0.6607	0.6715	0.1103	0.6618	0.0749	0.6647	0.0557
β_2	0.6607	0.6674	0.1157	0.6712	0.0788	0.6657	0.0550
δ_1	0.0000	-0.0018	0.1289	-0.0027	0.0671	-0.0023	0.0365
δ_2	0.0000	0.1544	1.0573	0.0637	0.5509	0.0294	0.1216
γ_0	0.3162	-0.0043	0.0957	0.0024	0.0662	-0.0041	0.0447
γ_1	0.3162	0.3197	0.1213	0.3212	0.0858	0.3181	0.0582
γ_2	0.3162	0.3219	0.1227	0.3243	0.0851	0.3226	0.0572
γ_3	0.3162	0.3271	0.1206	0.3168	0.0830	0.3182	0.0555
γ_4	0.3162	0.3235	0.1204	0.3218	0.0836	0.3188	0.0573
σ_U^2	2.7519	2.3698	1.1683	2.5371	0.8235	2.6265	0.5484
σ_V^2	1.0000	1.1145	0.3893	1.0618	0.2552	1.0273	0.1615
$\rho_{U,\eta}$	0.0000	0.0009	0.3552	0.0131	0.2672	-0.0034	0.2202
$\rho_{V,\eta}$	0.5000	0.5383	0.1838	0.5178	0.1187	0.5170	0.0823

TABLE 2. *Mean and Standard Errors of Estimators for simulation scheme 1*

		N = 250		N = 500		N = 1000	
	TRUE	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
β_0	0.0000	-0.0938	0.3075	-0.0457	0.1873	-0.0116	0.1158
β_1	0.6607	0.6611	0.1116	0.6608	0.0734	0.6606	0.0546
β_2	0.6607	0.6605	0.1163	0.6676	0.0790	0.6643	0.0554
δ_1	0.0000	-0.0028	0.1135	-0.0025	0.0707	-0.0011	0.0370
δ_2	0.0000	0.0346	1.1139	0.0507	0.4546	0.0146	0.1489
γ_0	0.3162	-0.0033	0.0961	0.0034	0.0660	-0.0031	0.0444
γ_1	0.3162	0.3193	0.1208	0.3200	0.0851	0.3182	0.0581
γ_2	0.3162	0.3218	0.1210	0.3239	0.0852	0.3224	0.0571
γ_3	0.3162	0.3271	0.1213	0.3165	0.0827	0.3185	0.0564
γ_4	0.3162	0.3264	0.1200	0.3232	0.0839	0.3193	0.0581
σ_U^2	2.7519	2.5362	1.2892	2.6001	0.9123	2.7050	0.6287
σ_V^2	1.0000	1.1028	0.4287	1.0687	0.2891	1.0155	0.1719
$\rho_{U,\eta}$	0.5000	0.4474	0.5394	0.4702	0.5196	0.4804	0.5028
$\rho_{V,\eta}$	0.5000	0.4934	0.2306	0.5047	0.1432	0.5059	0.0996

TABLE 3. *Mean and Standard Errors of Estimators for simulation scheme 2*

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 schemes the mean technical efficiency is equal to

$$E\left[\exp(-U)\right] = 0.3848.$$

We can appreciate how our estimator gives a plausible interval for the values of technical efficiency.

	$N = 250$		$N = 500$		$N = 1000$	
	$\rho_U = 0$	$\rho_U = 0.5$	$\rho_U = 0$	$\rho_U = 0.5$	$\rho_U = 0$	$\rho_U = 0.5$
Min.	0.001	0.000	0.001	0.000	0.001	0.000
1st Qu.	0.280	0.277	0.267	0.267	0.258	0.255
Median	0.439	0.441	0.417	0.418	0.404	0.405
Mean	0.428	0.425	0.402	0.401	0.391	0.390
3rd Qu.	0.566	0.570	0.540	0.541	0.529	0.529
Max.	1.000	1.000	1.000	1.000	1.000	0.828

Table 4. *Summary measures for the estimator of technical efficiency*

C. 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 results of the first-stage Probit regression.

Finally, we report here the results for the model where the production frontier does not s

The mean technical efficiency is reported at the bottom in Table 7, and Figure **??** provides the kernel density estimators of this measure for both the stochastic frontier model assuming exogeneity (dashed line) and the stochastic frontier model controlling for endogeneity (black line). The mean technical efficiency for the model controlling for endogeneity is low (0.61). However, estimates of technical efficiency vary considerably across farmers. Figure **??** shows that the distribution of the technical efficiency scores is bi-modal, with a group of farmers with high scores and the remaining farmers with low scores. As noted by Vo Hung Son et al. (1993), one potential explanation for this level of dispersion is that the assumption of common technology across farmers is violated. In particular, we observe that farmers participating in the soil conservation program tend to show high technical efficiency, whereas non-participants generally have low efficiency scores (see Table 8).

	Mean	St.Dev.	Min	Max
Output	1192.884	974.302	100.000	8468.200
Inputs				
Land	1.741	1.867	0.250	28.000
Labor	76.510	56.851	7.000	566.000
Fertilizers	206.485	160.995	0.000	1945.500
Pesticides	110.762	113.665	0.000	1116.600
Seeds	62.364	57.717	0.000	750.000
Slope	0.585	0.493	0.000	1.000
Environmental variables				
Organization	0.339	0.474	0.000	1.000
Education	2.058	0.994	0.000	4.706
Gender	0.932	0.252	0.000	1.000
Tenure	0.724	0.415	0.000	1.000
Off-Farm Income	2.151	3.129	0.000	8.666
Distance to plot (km)	0.791	1.015	0.000	7.570
Risk div	0.204	0.343	0.000	1.000
Participation				
Instruments	0.465	0.499	0.000	1.000
Distance to closest city (km)	1.056	0.838	0.050	5.100
Prop of families with electricity in canton	0.824	0.223	0.000	1.000
Average daily wage in canton	4.208	0.594	3.000	6.000

Table 5. Descriptive Statistics

	Estimate	Std. Error	z value	Pr(>|z|)
γ_0	-4.569	1.799	-2.540	0.011
γ_{Land}	0.364	0.153	2.381	0.017
γ_{Labour}	-0.317	0.160	-1.984	0.047
$\gamma_{Fertilizer}$	0.169	0.110	1.532	0.125
$\gamma_{Pesticides}$	-0.022	0.048	-0.469	0.639
γ_{Seeds}	0.099	0.065	1.508	0.132
$\gamma_{Organization}$	1.310	0.171	7.671	0.000
γ_{Educ}	0.121	0.074	1.625	0.104
γ_{Tenure}	0.430	0.170	2.538	0.011
$\gamma_{Off-farmincome}$	-0.001	0.022	-0.039	0.969
$\gamma_{Distance}$	-0.040	0.063	-0.633	0.527
γ_{Slope}	0.416	0.142	2.930	0.003
γ_{Div}	-0.257	0.230	-1.117	0.264
$\gamma_{DistCity}$	-1.469	0.715	-2.055	0.040
$\gamma_{DistCity^2}$	0.132	0.065	2.025	0.043
$\gamma_{FamElect}$	4.031	2.029	1.986	0.047
$\gamma_{FamElect \times DistCity}$	1.003	0.381	2.632	0.008
$\gamma_{DailyWage}$	0.585	0.374	1.566	0.117
$\gamma_{DailyWage \times FamElect}$	-0.712	0.471	-1.513	0.130
$\gamma_{DailyWage \times DistCity}$	0.305	0.156	1.962	0.050
$\gamma_{Organization \times DistCity}$	0.638	0.225	2.842	0.004
$\gamma_{Educ \times DistCity}$	-0.280	0.115	-2.430	0.015

TABLE 6. First Stage Probit Regression

	Exogeneity		Endogeneity	
	Estimate	Std. Err.	Estimate	Std. Err.
β_0	3.8502	0.1748	6.1286	0.2749
β_{Land}	0.368	0.0398	0.3767	0.0447
β_{Labour}	0.6161	0.0421	0.5628	0.0471
$\beta_{Fertilizer}$	0.0247	0.0218	0.0247	0.0219
$\beta_{Pesticides}$	-0.0112	0.0186	-0.0137	0.0198
β_{Seeds}	0.0531	0.0199	0.0485	0.0206
β_{Slope}	0.0593	0.0423	0.0304	0.0639
$\delta_{Organization}$			0.3038	0.2086
δ_{Educ}			-0.1479	0.0936
δ_{Tenure}			-0.1419	0.1821
$\delta_{Off-farmincome}$			0.0026	0.0307
$\delta_{Distance}$			-0.0537	0.1183
δ_{Slope}			0.0519	0.2179
δ_{Div}			0.3420	0.2347
$\delta_{Educ \times Participation}$			1.3127	0.8670
$\delta_{Participation}$			-5.2898	3.2339
$\rho_{U,\eta}$			-0.8121	0.0259
$\rho_{V,\eta}$			-0.6375	0.1360
σ_U^2			0.3053	0.2251
σ_V^2	0.1846	0.012	0.2048	0.0259
Mean TE	0.0523		0.6100	

TABLE 7. *Estimation of the efficiency frontier with and without accounting for treatment endogeneity.*

	Estimate	Std. Error	T-stat	P-value
Intercept	0.402	0.007	59.971	0.000
Participation	0.446	0.012	35.987	0.000

Table 8. Regression of TE on participation dummy