ORIGINAL PAPER

Endogeneity in stochastic frontier models with 'wrong' skewness: copula approach without external instruments

Rouven E. Haschka[1,2]

Abstract

Stochastic frontier models commonly assume positive skewness for the inefficiency term. However, when this assumption is violated, efficiency scores converge to unity. The potential endogeneity of model regressors introduces another empirical challenge, impeding the identification of causal relationships. This paper tackles these issues by employing an instrument-free estimation method that extends joint estimation through copulas to handle endogenous regressors and skewness issues. The method relies on the Gaussian copula function to capture dependence between endogenous regressors and composite errors with a simultaneous consideration of positively or negatively skewed inefficiency. Model parameters are estimated through maximum likelihood, and Monte Carlo simulations are employed to evaluate the performance of the proposed estimation procedures in finite samples. This research contributes to the stochastic frontier models and production economics literature by presenting a flexible and parsimonious method capable of addressing wrong skewness of inefficiency and endogenous regressors simultaneously. The applicability of the method is demonstrated through an empirical example.

✉ Rouven E. Haschka
haschka@zu.de; rouven.haschka@uni-corvinus.hu

1 Chair of Business Analytics & Data Science, Zeppelin University, Am Seemoser Horn 20, 88045 Friedrichshafen, Germany

2 Institute of Strategy and Management, Corvinus University, Fővám tér 8, 1093 Budapest, Hungary

1 Introduction

The classical assumption in stochastic frontier (SF) production models is that inefficiency exhibits positive skewness, resulting in the composite error, i.e., the regression residuals, having negative skewness. However, in empirical applications, the residuals may exhibit positive skewness, which contradicts the assumption of positively skewed inefficiency. Waldman (1982) first demonstrated that if the residuals from the SF model have 'wrong' skewness, i.e., positive, inefficiency variance is biased towards zero. Consequently, efficiency scores tend to be one, leading to false conclusions of high efficiency (Hafner et al. 2018). Green and Mayes (1991) argue that this either indicates 'super efficiency' (all firms in the industry operate close to the frontier) or the inappropriateness of the SF analysis technique to measure inefficiencies. Thus, implausibly high efficiency scores obtained under the classical SF specification can indicate misspecification of inefficiency skewness (Haschka and Wied 2022).

Another significant empirical challenge arises in the case of regressor endogeneity. Endogeneity can be loosely defined as dependence between regressors and errors, which is particularly important for SF models, as this dependence may stem from inefficiency, idiosyncratic noise, or both (Tran and Tsionas 2015). Linking endogenous covariate information with composite errors can lead to biased estimates for causal effects if the methods used assume regressor exogeneity (Haschka and Herwartz 2022, 2020). The standard approach to handling the endogeneity problem in SF models is to use likelihood-based instrumental variable (IV) estimation methods (Amsler et al. 2016; Kutlu 2010; Prokhorov et al. 2020; Tran and Tsionas 2013). However, a general drawback of such methods is their reliance on the availability of external information to construct instruments. Instruments, if available at all, are often subject to potential pitfalls if they fail to adequately meet the two required conditions: they must be sufficiently correlated with the endogenous regressors and uncorrelated with the composite errors term. Thus, a potential difficulty in implementing IV-based estimators arises when there is no external information available to construct appropriate instruments (Haschka et al. 2020).

Given that suitable instrumental information in SF models is often scant, unavailable, or weak, this study proposes an SF model with a data-driven choice of correct or wrong skewness, which we conceptualize as an extension of the IV-free joint estimation model using copulas introduced by Park and Gupta (2012). Copula techniques have been successfully adopted for classical SF models (Tran and Tsionas 2015), and we extend them to address cases of wrong skewness. The method relies on a copula function to directly model dependencies between endogenous regressors and composite errors, thus eliminating the need for external information. Specifically, copulas allow for the separate modeling of marginal distributions of endogenous regressors and composite errors, while capturing their dependency. We construct the joint distribution of the endogenous regressor and composite error, which accommodates mutual dependency between them. Subsequently, we use this joint distribution to derive the likelihood function,

which distinguishes between correct and wrong skewness through an indicator function, and maximize it to obtain consistent estimates of model parameters.

The empirical analysis aims to impartially explore the determinants of firm performance based on data from 16,641 Vietnamese firms in 2015. A key observation is that the presence of wrong skewness is compounded by endogeneity, while the suggested estimator provides an unbiased perspective. The empirical findings indicating wrong skewness imply a growing number of inefficient firms persisting in the market, challenging the assumption of a competitive landscape. This persistence in inefficiency is likely influenced by factors such as corruption and the constraints imposed by the communist regime, hindering the establishment of liberal and competitive market conditions. Consequently, policy interventions are deemed necessary to incentivize firms to optimize their processes and improve efficiency.

The paper is organized as follows. Section 2 presents the model and discusses the copula approach to deal with regressor endogeneity in SF models with wrong skewness. In Sect. 3, we examine the finite sample performance of the proposed approach through Monte Carlo simulations. An empirical application is provided in Sect. 4. We revisit the methodology in Sects. 5, and 6 concludes with a summary of our contribution.

2 Methodology

Consider the standard SF model given by:

$$y_i = x_i'\beta + z_i'\delta + \underbrace{v_i - u_i}_{e_i}, \quad i = 1, \dots, n, \tag{1}$$

Here, y_i represents the output of producer i, x_i is an $L \times 1$ vector of exogenous inputs, z_i is a $K \times 1$ vector of endogenous inputs, β and δ are $L \times 1$ and $K \times 1$ vectors of unknown parameters to be estimated, respectively. Additionally, v_i is a symmetric random error, u_i is a one-sided random disturbance representing technical efficiency, and the composite error is denoted as $e_i = v_i - u_i$. It is assumed that x_i is uncorrelated with e_i, while z_i is allowed to be correlated with e_i, giving rise to endogeneity, but we make no assumption regarding the source of endogeneity, being it correlation with v_i or with u_i. Furthermore, it is assumed that u_i and v_i are independent, and the skewness of u_i is left unrestricted. This discussion can be readily extended to cases where the (exogenous) environmental variables are included in the distribution of u_i (Battese and Coelli 1995; Haschka and Herwartz 2022).

2.1 'Wrong' skewness of inefficiency distribution

Adhering to standard SF practices, we presume that $v_i \sim \mathrm{N}(0, \sigma_v^2)$ represents symmetric, two-sided idiosyncratic noise. In this context, our model operates under the assumption that skewness issues arise from the inefficiency term, while v_i itself is

symmetric.[1] Following the approach of Hafner et al. (2018), we delineate two cases for u_i that characterise the distributional shape:

'Correct' skewness:

$$u_i \sim N_{[0,\infty)}(0, \gamma^2), \quad \gamma > 0 \tag{2}$$

'Wrong' skewness:

$$u_i \sim N_{[0,a_0|\gamma|)}(a_0|\gamma|, \gamma^2), \quad \gamma < 0 \tag{3}$$

The assumption in (2) is widely acknowledged, characterising the correct skewness of u_i (and consequently e_i) due to the strictly decreasing density of u_i in the interval $[0, \infty)$ (Kumbhakar and Lovell 2003), and constitutes the classical SF model. On the contrary, wrong skewness is induced by (3), where $a_0 \approx 1.389$ represents the non-trivial solution of $\frac{\phi(0)}{\Phi(0)} = a_0 + \frac{\phi(a_0)-\phi(0)}{\Phi(a_0)-\Phi(0)}$, and the density of u_i is strictly increasing and bounded in $[0, a_0|\gamma|]$.

It is noteworthy that the expectations of both u_i and e_i remain unaffected by the skewness' sign (Hafner et al. 2018). Thus, the inefficiency variance and the sign of skewness are directly linked because $\gamma > 0$ ($\gamma < 0$) induces correct (wrong) skewness, but $\mathbb{E}[u_i]$ and $\mathbb{E}[e_i]$ are not influenced by the sign of γ. By convolution of v with u and integration, the density of $e_i = v_i - u_i$ obtains as:

'Correct' skewness:

$$g_e^+(e) = \frac{2}{\sigma}\phi\left(\frac{e}{\sigma}\right)\Phi\left(-\frac{e\gamma}{\sigma\sigma_v}\right), \tag{4}$$

'Wrong' skewness:

$$g_e^-(e) = \frac{1}{\sigma\left(\Phi(a_0) - \Phi(0)\right)}\phi\left(\frac{e - a_0\gamma}{\sigma}\right)\left[\Phi\left(A_w + \frac{a_0\sigma}{\sigma_v}\right) - \Phi(A_w)\right],$$
$$A_w = \frac{e - a_0\gamma}{\sigma}\frac{\gamma}{\sigma_v}, \tag{5}$$

with $\sigma^2 = \gamma^2 + \sigma_v^2$ and $\int e g_e^+(e)\, de = \int e g_e^-(e)\, de$. It is important to note that under correct skewness, e follows a (1x1)-dimensional closed skew normal (CSN) distribution, denoted as $e^+ \sim CSN(0, \sigma^2, -\frac{\gamma}{\sigma_v\sigma}, 0, 1)$, while, as demonstrated by Haschka and Wied (2022), under wrong skewness, e follows a (1x2)-dimensional CSN distribution, denoted as $e^- \sim CSN_{1\times2}\left(a_0\gamma, \sigma^2, \begin{pmatrix} \gamma/\sigma \\ -\gamma/\sigma \end{pmatrix}, \begin{pmatrix} -a_0\sigma \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}\right)$. Con-

---

[1] Within the SF literature, there are discussions on various sources of wrong skewness. While it is commonly attributed to the inefficiency term (as in our case), asymmetry of idiosyncratic noise (Badunenko and Henderson 2023; Bonanno et al. 2017; Horrace et al. 2023; Son et al. 1993), or dependence between noise inefficiency and idiosyncratic noise (Smith 2008; Bonanno et al. 2017; Bonanno and Domma 2022) is also considered.
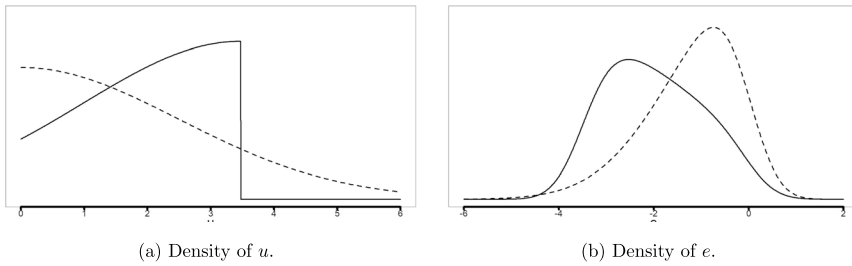
(a) Density of u.   (b) Density of e.

Fig. 1 Densities of u and e for $\gamma = 2.5$, i.e., correct skewness (dotted lines) and $\gamma = -2.5$, i.e., wrong skewness (solid lines); with $\sigma_v = .5$

sequently, it becomes evident that another advantage of the distribution in (3) is its association with the well-known CSN distribution. As a result, properties of the CSN distribution, such as moments, can be derived (Flecher et al. 2009).

Although Li (1996) argues that one-sided distributions with an unbounded range always exhibit positive skewness, Johnson et al. (1995) demonstrates that the two-parameter Weibull distribution can have both positive and (small) negative skewness for specific parameter combinations. Accordingly, this distribution could be an alternative without needing to impose an upper boundary. However, since the upper bound of the distribution in (3) depends on γ, the advantage of using the truncated normal distribution in (3) is that it entails negative skewness without necessitating the identification of additional parameters. Moreover, other parsimonious distributions for u that can produce wrong skewness, such as the negative skew-exponential distribution discussed in Hafner et al. (2018), may also be considered. However, in such cases, it would not be clear if the distribution of e will be known or has a closed-form expression.

The configuration of the inefficiency distribution under both correct and wrong skewness is illustrated in Panel (a) of Fig. 1, with the corresponding distributions of composite errors presented in Panel (b). When $\gamma > 0$, the distribution of u exhibits positive skewness, while the distribution of e has negative skewness. Conversely, for $\gamma < 0$, the skewness of u becomes negative, and that of e is positive. It is important to emphasize that the adopted one-sided distribution for inefficiency, which can result in both correct and wrong skewness, is parsimonious. Alternative approaches allowing for a data-driven selection of either correct or wrong skewness often involve identifying multiple parameters determining the inefficiency distribution (see, e.g., Tsionas 2017, for Weibull inefficiency), may not result in a well-known distribution for composite errors, or require a priori determination of the skewness sign (corrected OLS or modified OLS). However, such determinations are challenging in the presence of endogeneity.

2.2 Joint estimation using copulas

Consider $F(z_1, \ldots, z_K, e)$ and $f(z_1, \ldots, z_K, e)$ as the joint distribution and joint density of endogenous regressors $(z_1, \ldots, z_K)$ and composite errors e, respectively. In

practical applications, $F(\cdot)$ and $f(\cdot)$ are unknown due to the unobservability of e and thus require estimation. In line with the methodology proposed by Park and Gupta (2012), we employ a copula approach to approximate this joint density. The copula serves as a tool to capture dependence within the joint distribution of endogenous regressors and composite errors.

Let $\boldsymbol{\omega}_{z,i} = (F_{z1}(z_{1i}), \ldots, F_{zK}(z_{Ki}))'$ and $\omega_{e,i} = G(e_i; \sigma_v, \gamma)$ represent the margins $(\boldsymbol{\omega}_{z,i}, \omega_{e,i})' \in [0,1]^{K+1}$ based on a probability integral transform. Here, the F's signify the respective marginal cumulative distribution functions of the observed endogenous regressors, and $G(e_i; \sigma_v, \gamma)$ is the cumulative distribution function of the CSN distribution for errors, subject to the sign of γ. Drawing on the approach by Haschka (2021) and Tran and Tsionas (2015), we substitute $F_1(z_{1i}), \ldots, F_p(z_{pi})$ with their respective empirical counterparts in a first stage. Given observed samples of z_{ji}, $j = 1, \ldots, p$; $i = 1, \ldots, n$, we utilise the empirical cumulative distribution function (ecdf) of z_j, denoted as $\hat{F}_j = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(z_{ji} \leq z_{0j})$.

Utilising a Gaussian copula, $\hat{\boldsymbol{\xi}}_{z,i} = (\Phi^{-1}(\hat{F}_{z1}(z_{1i})), \ldots, \Phi^{-1}(\hat{F}_{zK}(zKi)))'$, and $\hat{\xi}_{e,i} = \Phi^{-1}(G(\hat{e}_i; \hat{\sigma}_v, \hat{\gamma}))$, both follow a standard multivariate normal distribution with a dimension of $(K+1)$ and a correlation matrix Ξ. Subsequently, the joint density can be expressed as

$$
\begin{aligned}
f(z_i, e_i) = & \frac{1}{\sqrt{\det(\Xi)}} \exp\left(-\frac{1}{2} \begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\xi}_{e,i} \end{pmatrix}' (\Xi^{-1} - I) \begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\xi}_{e,i} \end{pmatrix} \right) \\
& \times g(e_i; \sigma_v, \gamma) \times \prod_{k=1}^{K} f_{zk}(\check{z}_{ki}),
\end{aligned}
\tag{6}
$$

where $\xi_{e,i}$ and $g(e_i; \sigma_v, \gamma)$ are again subject to either correct or wrong skewness. The copula density in the first row establishes a connection between the error and all endogenous variables. Meanwhile, the densities in the second row describe the marginal behaviour. The marginal densities $f_{zk}(\check{z}_{ki})$ in (6) do not involve any parameters of interest and can be omitted when deriving the likelihood since they function as normalising constants.

To simultaneously determine the choice of correct or wrong skewness based on the sign of γ, we adopt the approach of Hafner et al. (2018) and incorporate an indicator function into the likelihood. While Hafner et al. (2018) propose choosing correct or wrong skewness a priori by examining the skewness of the OLS residual, our method estimates the sign of γ simultaneously with all other parameters. This is preferred because any pre-determination of residual skewness could be influenced by (potential) endogeneity. Consequently, the likelihood function is given by:

$$L(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}) \propto \mathbb{1}(\gamma > 0) \prod_{i=1}^{n} \frac{1}{\sqrt{\det(\Xi)}}$$

$$\exp\left(-\frac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{+} \end{pmatrix}' (\Xi^{-1} - I)\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{+} \end{pmatrix}\right) g^{+}(e_i; \sigma_v, \gamma)$$

$$+ \mathbb{1}(\gamma < 0) \prod_{i=1}^{n} \frac{1}{\sqrt{\det(\Xi)}} \exp\left(-\frac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{-} \end{pmatrix}' (\Xi^{-1} - I)\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{-} \end{pmatrix}\right) g^{-}(e_i; \sigma_v, \gamma)$$

$$+ \mathbb{1}(\gamma = 0) \prod_{i=1}^{n} \frac{1}{\sqrt{\det(\Xi)}} \exp\left(-\frac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{0} \end{pmatrix}' (\Xi^{-1} - I)\begin{pmatrix} \hat{\boldsymbol{\xi}}_{z,i} \\ \hat{\boldsymbol{\xi}}_{e,i}^{0} \end{pmatrix}\right) \phi(e_i; \sigma_v).$$

$$(7)$$

To explicitly account for the scenario of only fully efficient firms, the likelihood also accommodates $\gamma = 0$.[2] In this case, the marginal distribution of the errors follows a normal distribution with mean zero and variance σ_v^2, represented as $\xi_{e,i}^0 = e_i / \sigma_v$. Notably, our approach encompasses those of Hafner et al. (2018), Tran and Tsionas (2015), and Park and Gupta (2012). Specifically, under the assumption of exogeneity of all regressors ($\Xi = I$), the likelihood in (7) reduces to that in Hafner et al. (2018). In the case of correct skewness ($\gamma > 0$), corresponding to the traditional SF model, the likelihood collapses to that in Tran and Tsionas (2015). Finally, for the scenario of only fully efficient firms ($\gamma = 0$), it reduces to that in Park and Gupta (2012). Subsequently, the likelihood is logarithms and maximised with respect to the vector of unknown parameters $\boldsymbol{\theta} = (\beta, \delta, \sigma^2 v, \gamma, \text{vechl}[\Xi])$, where $\text{vechl}[\Xi] = (\rho_1, \dots, \rho_K)'$ stacks the lower diagonal elements of the correlation matrix Ξ into a column vector.

With parameter estimates, technical inefficiency u_i can be predicted using Jondrow et al. (1982) as follows:

'Correct' skewness:

$$\hat{u}_i = \hat{E}(u_i \mid e_i) = \frac{\sqrt{\hat{\gamma}^2 + \hat{\sigma}_v^2}(\hat{\gamma}/\hat{\sigma}_v)}{1 + (\hat{\gamma}/\hat{\sigma}_v)^2} \left[\frac{\phi\left((\hat{\gamma}/\hat{\sigma}_v)\hat{e}_i / \sqrt{\hat{\gamma} + \hat{\sigma}_v}\right)}{1 - \Phi\left((\hat{\gamma}/\hat{\sigma}_v)\hat{e}_i / \sqrt{\hat{\gamma} + \hat{\sigma}_v}\right)} \right.$$

$$\left. - \frac{(\hat{\gamma}/\hat{\sigma}_v)\hat{e}_i}{\sqrt{\hat{\gamma}^2 + \hat{\sigma}_v^2)}} \right] \qquad (8)$$

'Wrong' skewness:

$$\hat{u}_i = \hat{E}(u_i \mid e_i) = \int_0^{a_0|\hat{\gamma}|} \exp\{-u_i\} f^{-}(u_i \mid e_i) du_i \qquad (9)$$

---

[2] In case of exogeneity, the likelihood function is continuous in γ, as shown in Appendix A.1 of Hafner et al. (2018). Intuitively, this should also hold for nonzero diagonal elements in Ξ.

Note that in the case of correct skewness, the predicted technical inefficiency has a closed-form expression, while under wrong skewness, the integral needs to be solved numerically (Hafner et al. 2018).

2.2.1 Model identifiability

In our framework, model identification hinges on two key conditions: (i) the distribution of endogenous regressors must differ from that of the composite error (for an in-depth discussion on identification in copula-based endogeneity-correction models, see Haschka 2022b; Papadopoulos 2022; Park and Gupta 2012). Consequently, the model maintains identification as long as γ is not zero (or very close to zero), and endogenous regressors deviate from a normal distribution. However, identification collapses if $\gamma = 0$ (resulting in a normal composite error, in which case our model aligns with that of Park and Gupta 2012), and endogenous regressors follow a normal distribution. In such a scenario, the joint distribution of endogenous regressors and the composite error becomes multivariate normal, implying a *linear* relationship between the endogenous regressors and the composite error. Consequently, we would be unable to distinguish the *linear* effect of the endogenous regressor on the outcome. To address this, a formal separation of $\mathbb{E}[y|z]$ from $\mathbb{E}[e|z]$ without IV information is impossible if both are linear functions, as indicated by joint normality (Haschka 2022b).

In contrast, when the joint distribution is not multivariate normal (with $\gamma = 0$, indicating the non-normality of endogenous regressors), the relationship between regressors and errors becomes non-linear. Identification in this context does not require IVs and can be accomplished through the joint distribution of e and z. Although $\mathbb{E}[y|z]$ remains a linear function due to the specification of a linear regression model, non-normality implies that $\mathbb{E}[e|z]$ becomes a non-linear function. This non-linearity facilitates the separation of variation attributed to endogenous regressors from the variation due to the composite error, as discussed in Tran and Tsionas (2015). Consequently, in empirical applications, it is essential to assess the marginal distribution of endogenous regressors before estimation-an approach commonly adopted in empirical literature utilizing copula-based identification (e.g., Haschka 2022b; Papadopoulos 2022; Park and Gupta 2012).

Through the utilization of a Gaussian copula, model identification also requires (ii) a linear dependence between ξz and ξe, enabling the correlation to be expressed through pairwise Pearson coefficients. Consequently, the model exclusively accommodates (iii) continuous endogenous regressors (or discrete with numerous distinct outcomes). In instances where z_k is continuous, $\xi_{z,k} = (\Phi^{-1}(F_{zk}(z_k)))$ follows a standard normal distribution. Conversely, if z_k is binary, $\xi_{z,k}$ would also be binary, and multivariate normality with ξ_e would not be guaranteed. While the continuity of z_k is crucial and its verification is straightforward, confirming the assumption of Gaussian-type dependence is more challenging empirically due to the unobservability of errors. However, since any copula capable of modeling multivariate dependency structures can be employed, this identification assumption can be readily substituted when opting for a different copula. Despite this, in cases where the true dependence deviates from the Gaussian copula assumption, existing literature has demonstrated

the robustness of the Gaussian copula in flexibly capturing various non-Gaussian dependencies (Becker et al. 2021; Haschka 2022b; Park and Gupta 2012). Papadopoulos (2022) suggest testing for the multivariate normality of $\boldsymbol{\xi}_z$ to assess the suitability of the Gaussian copula, allowing for the determination of whether the assumption holds for a specific part of the model. Nevertheless, given the Gaussian copula's robustness in modeling diverse dependency structures, significance in this context does not necessarily imply adverse effects on model identification.

2.2.2 Bootstrap inference and asymptotic behaviour

Copula-based endogeneity-correction models are commonly formulated as two-step estimation approaches, with the marginal cumulative distribution function of endogenous regressors determined a priori through a data-driven process.[3] In the initial stage, we derive the margins $\hat{\boldsymbol{\omega}}_{z,i}$, representing the probability integral transformed endogenous variables through the empirical cumulative distribution function. This serves as an estimator for the cumulative distribution function (CDF) (Joe and Xu 1996). In the subsequent step, these estimated margins are incorporated into the likelihood function in (7) as plug-in estimates. This introduces two significant implications for the estimator in the second stage. Firstly, the estimator cannot achieve unbiasedness since estimated CDFs from the sample are used instead of those from the population (Genest et al. 1995). Therefore, it is crucial to establish consistency, a validation that will be undertaken through Monte Carlo simulations. Secondly, standard errors in the second stage are generally unreliable as uncertainties from the first stage are not taken into account.

The employed copula estimator can be categorized as a semiparametric method since it combines the nonparametric distribution of endogenous regressors from the first stage with the parametric CSN distribution of the composite errors and the parametric Gaussian copula constituting the likelihood in the second stage (Genest et al. 1995). In semiparametric models, deriving Hessian-based asymptotic standard errors can be notably challenging, making bootstrap procedures a natural alternative for evaluating estimation uncertainty (Haschka and Herwartz 2022; Park and Gupta 2012). Consequently, we employ bootstrapping to derive standard errors and confidence intervals.

Recently, Breitung et al. (2023) argued that it remains uncertain a priori whether the standard properties of maximum likelihood (ML) estimation hold for IV-free copula-based endogeneity corrections and under which assumptions they may be applicable. The challenge of formulating precise statements about limiting properties in the presence of a nonparametrically generated regressor is a highly intricate task. This challenge is inherent in all copula-based approaches that rely on the joint estimation of errors and endogenous regressors with a priori estimated cumulative distribution

---

[3] There are a few exceptions to this standard practice. Tran and Tsionas (2021) employs sieve maximum likelihood estimation to simultaneously estimate marginal cumulative distribution functions, while Haschka (2022a, 2023) do the same within a Bayesian framework using Dirichlet priors on explanatory variables.

functions (Haschka 2022b; Park and Gupta 2012; Tran and Tsionas 2015). Nevertheless, Papadopoulos (2022) and Tran and Tsionas (2015) conjectured that such estimators exhibit consistency and asymptotic normality in stochastic frontier (SF) settings (for additional simulation-based evidence on asymptotic behavior, see Haschka 2022b; Haschka and Herwartz 2022). Although specific consistency (i.e., robustness) theory for quasi-maximum likelihood estimation (quasi-MLE) for Gaussian copula-based models exists (Prokhorov and Schmidt 2009), it is unclear whether these results are valid in the case of generated regressors. However, some insights from related literature (Breitung et al. 2023; Genest et al. 1995) can be applied. Under exogeneity, the generated regressors do not impact the asymptotic distributions, and the estimator asymptotically follows a standard normal distribution (Breitung et al. 2023). In the case of endogeneity, the limiting distribution is unknown. As demonstrated by Breitung et al. (2023) for linear regression models with Gaussian outcomes (i.e., no SF specifications), the asymptotic distribution depends on unknown parameters (under endogeneity). In this context, we assume that in the case of an SF specification (non-zero γ), the asymptotic distribution differs from that derived in Breitung et al. (2023) - more precisely, it depends on other unknown parameters. Consequently, no general theoretical statement can be made about the asymptotic behavior of the estimator.

3 Monte Carlo simulations

To assess the finite sample performance of the proposed copula estimator, we conduct Monte Carlo simulations based on the following data generating process (DGP):

$$y_i = \beta x_i + \delta z_i + v_i - u_i \tag{10}$$

$$z_i = \alpha s_i + \eta_i \tag{11}$$

$$u_i \sim \begin{cases} \mathrm{N}_{[0,\infty)}(0, \gamma^2) & \text{if } \gamma > 0, \\ \mathrm{N}_{[0,a_0|\gamma|]}(a_0|\gamma|, \gamma^2) & \text{if } \gamma < 0, \end{cases} \tag{12}$$

In (12), positive (negative) values of γ result in the distribution of u_{it} having positive (negative) skewness, with both distributions having the same expectation. The random variables x_i and s_i are each generated independently as $\chi^2(2)$. To introduce endogeneity, the vector of errors $(v_i, \eta_i)'$ is generated by:

$$\begin{pmatrix} v_i \\ \eta_i \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \tag{13}$$

In our experiments, we fix $\alpha = \beta = \delta = .5$ and rescale $v_i = v_i/2$ to achieve a signal-to-noise ratio of $\gamma/\sigma_v = 2$, which is common is assessing performance of SF models in Monte Carlo simulations (see e.g., Chen et al. 2014; Haschka and Wied 2022). To assess scenarios under exogeneity and endogeneity, we set $\rho = \{0, .35, .7\}$. While the scenario with exogeneity is intended to highlight the costs of accounting for endogeneity, introducing endogeneity in a stepwise manner allows monitoring

how the performance is affected. We further distinguish $\gamma = \{-1, 1\}$ to introduce either negative or positive skewness. Finally, the sample size is $N = 750$, and each experiment is replicated 1000 times. For comparison purposes, we also compute the Maximum Likelihood Estimator (MLE) by Hafner et al. (2018), which considers correct and wrong skewness but does not account for endogeneity, an instrument-based Generalised Method of Moments (GMM) estimator (Tran and Tsionas 2013), and an IV-free copula estimator (Tran and Tsionas 2015), both of which address endogenous regressors but assume correctly skewed inefficiency.

Estimation results for scenarios of exogeneity and endogeneity with either correct or wrong skewness are shown in Table 1. When the skewness of the inefficiency distribution is correctly specified (upper panel) and the regressors are exogenous, all estimators are unbiased. In this scenario, ML is the method of choice as it yields the smallest standard deviations, while the remaining approaches (GMM, copula, proposed) are characterized by higher uncertainty as they unnecessarily allow for endogeneity. This scenario thus highlights the costs of incorrectly accounting for endogeneity.

Under moderate endogeneity ($\rho = .35$), ML deteriorates and exhibits bias for all model parameters. This bias becomes even more pronounced when endogeneity becomes stronger ($\rho = .7$). While GMM, copula, and the proposed estimator perform equally well and remain unbiased regardless of the degree of endogeneity, estimates obtained by GMM are most tightly centered around true values, primarily because it is IV-based and access to valid instruments strongly benefits efficiency (for similar findings, see Tran and Tsionas 2015).

Misspecifying the skewness of the inefficiency distribution (lower panel) has particularly detrimental effects on the GMM and copula estimators. Even in the absence of endogeneity, both estimators yield biased slope coefficients. Misspecification of the error distribution due to wrong skewness is falsely attributed to endogeneity, leading to biased estimates (for further simulation-based evidence, see Haschka and Herwartz 2022). Consequently, efficiency scores are also biased. When introducing endogeneity, GMM and copula remain biased due to misspecified error distribution.

Moreover, the ML estimator yields an incorrect sign of skewness when the true inefficiency exhibits wrong skewness. While the ML estimator is capable of detecting wrong skewness under exogeneity, it fails to do so under endogeneity. Specifically, even with moderate endogeneity ($\rho = .35$), if the true coefficient is $\gamma = -1$, the ML estimator yields a coefficient estimate of .1892, incorrectly suggesting correct skewness. This highlights a substantial hindrance in detecting wrong skewness in the presence of endogenous regressors. By contrast, the proposed approach remains generally unaffected, is unbiased for all coefficients, and provides accurate assessments of efficiency.

4 Application

We demonstrate the practicality of the proposed approach through an empirical analysis, assessing firm efficiency using data from Vietnamese firms in the year 2015 obtained from the Vietnam Enterprise Survey (VES). The VES, conducted annually

Table 1 Monte Carlo simulation results are presented for simulations based on the DGP in Eqs. (10) to (13) under scenarios of exogeneity ($\rho = 0$) and endogeneity ($\rho = .35, .7$), considering both correct skewness ($\gamma > 1$, upper panel) and wrong skewness ($\gamma < 0$, lower panel)

		Parameter	ML (Hafner et al. 2018)			GMM (Tran and Tsionas 2013)			Copula (Tran and Tsionas 2015)			Proposed		
			Mean	Sd	MSE	Mean	Sd	MSE	Mean	Sd	MSE	Mean	Sd	MSE
$\gamma = 1$	$\rho = 0$	β	.5001	.0211	.0004	.4999	.0220	.0005	.4999	.0204	.0004	.5018	.0215	.0005
		δ	.5003	.0304	.0009	.4998	.0998	.0099	.4859	.1442	.0209	.4952	.1438	.0207
		σ_v	.4967	.0620	.0038	.4977	.0629	.0038	.4958	.0615	.0038	.5011	.0785	.0063
		γ	.9966	.1044	.0110	.9967	.1481	.0220	.9834	.1822	.0334	.9836	.1841	.0342
		ρ	–	–	–	.0099	.1627	.0264	.0192	.1822	.0336	-.0074	.1821	.0332
		$\mathbb{E}[u|e]$	–	–	.0672	–	–	.0688	–	–	.0691	–	–	.0701
	$\rho = .35$	β	.5003	.0201	.0004	.5002	.0220	.0005	.4992	.0208	.0004	.4996	.0218	.0005
		δ	.6401	.0300	.0205	.5003	.0998	.0099	.4985	.1449	.0210	.4989	.1440	.0207
		σ_v	.3918	.0608	.0154	.5091	.0624	.0039	.5194	.0619	.0042	.4981	.1414	.0120
		γ	1.134	.0866	.0255	1.093	.1484	.0306	1.013	.1830	.0336	1.034	.1850	.0354
		ρ	–	–	–	.3511	.1647	.0271	.3481	.1850	.0342	.3477	.1849	.0342
		$\mathbb{E}[u|e]$	–	–	.2391	–	–	.0692	–	–	.0696	–	–	.0713
	$\rho = .7$	β	.4981	.0179	.0004	.5058	.0222	.0005	.5001	.0221	.0005	.4971	.0222	.0005
		δ	.7478	.0296	.0622	.4950	.0999	.0100	.5032	.1457	.0212	.5037	.1450	.0210
		σ_v	.2560	.0501	.0619	.4980	.0620	.0038	.4942	.0632	.0040	.5016	.0801	.0067
		γ	1.311	.0710	.0980	.9983	.1488	.0219	.9909	.1842	.0344	.9879	.1861	.0348
		ρ	–	–	–	.6920	.1661	.0278	.7029	.1885	.0356	.6981	.1871	.0350
		$\mathbb{E}[u|e]$	–	–	.4299	–	–	.0699	–	–	.0695	–	–	.0722

Table 1 (continued)

Parameter		ML (Hafner et al. 2018)			GMM (Tran and Tsionas 2013)			Copula (Tran and Tsionas 2015)			Proposed		
		Mean	Sd	MSE	Mean	Sd	MSE	Mean	Sd	MSE	Mean	Sd	MSE
$\gamma = -1$	$\rho = 0$												
	β	.5000	.0199	.0004	.4996	.0208	.0004	.4979	.0200	.0004	.5012	.0201	.0004
	δ	.4985	.0275	.0007	.3895	.0903	.0203	.4011	.1378	.0287	.4931	.1759	.0310
	σ_v	.5018	.0304	.0013	.5011	.0298	.0014	.4891	.0302	.0016	.5035	.0533	.0041
	γ	-.9719	.0790	.0070	.7794	.1312	3.182	.7395	.1695	3.053	-.9637	.2351	.0566
	ρ	–	–	–	.1411	.1645	.0469	.1240	.1773	.0469	-.0055	.2422	.0587
	$\mathbb{E}[u|e]$	–	–	.0590	–	–	.6922	–	–	.7043	–	–	.0761
$\rho = .35$	β	.4993	.0198	.0004	.5091	.0219	.0005	.5019	.0197	.0004	.5006	.0200	.0004
	δ	.6615	.0298	.0270	.2874	.1184	.0592	.2970	.1550	.0652	.5150	.1934	.0376
	σ_v	.4591	.0320	.0027	.4817	.0296	.0012	.4594	.0050	.0017	.5162	.1885	.0358
	γ	.1892	.0755	1.420	.7618	.1351	3.122	.6015	.2084	2.608	-.9901	.2338	.0548
	ρ	–	–	–	.5291	.1610	.0580	.5406	.1841	.2245	.3515	.2448	.0599
	$\mathbb{E}[u|e]$	–	–	.3094	–	–	.7194	–	–	.7522	–	–	.0766
$\rho = .7$	β	.4995	.0186	.0003	.4990	.0178	.0003	.4993	.0184	.0003	.5019	.0193	.0004
	δ	.7469	.0281	.0619	.1411	.1422	.7583	.1400	.1800	.0619	.5161	.1918	.0371
	σ_v	.4140	.0327	.0085	.4275	.0302	.0359	.4140	.0321	.0085	.5011	.0337	.0013
	γ	.3615	.0718	1.859	.7318	.1388	3.139	.4166	.2430	2.059	-.9828	.2299	.0532
	ρ	–	–	–	.9242	.1590	.0755	.9747	.1910	.1119	.7031	.2314	.0537
	$\mathbb{E}[u|e]$	–	–	.3676	–	–	.7271	–	–	.8292	–	–	.0750

The simulations involve a sample size of $N = 750$. The table displays the mean estimates, standard deviations of the respective coefficients, and mean squared errors

Table 2 Summary Statistics: All variables are measured in million Vietnamese Dong

Variable	N. obs	Mean	Median	Std. Dev	IQR	Min	Max
Revenues	16,641	40,280	2659	471,973	9260	1	39,175,496
Wages	16,641	3354	466	22,561	970	3	1,425,235
Assets	16,641	16,799	1607	160,314	3525	3	8,356,849

Revenues represent the total revenue from sales and services, wages indicate the total expenditure on wages, and assets denote the total value of fixed assets

by the General Statistics Office (GSO) of Vietnam, is a nationally representative survey encompassing all firms with 30 or more employees, along with a representative sample of smaller firms (O'Toole and Newman 2017). Our evaluation of firm performance follows established approaches in the literature, examining the interplay between firm revenue, wages, and assets (e.g., Haschka et al. 2021, 2023). Descriptive statistics for the involved variables are presented in Table 2.

By treating each firm as an individual producer, we adopt a log-linear Cobb–Douglas production function and specify the following model:

$$\log revenue_i = \alpha + \delta_1 \log wages_i + \delta_2 \log assets_i + v_i - u_i, \tag{14}$$

where $i = 1, \dots, 16,474$ represents firms, and $v_i \sim N(0, \sigma_v^2)$ is idiosyncratic noise. Our specification differs from related models in the following two aspects. First, we allow stochastic inefficiency to vary across i and consider both correct and wrong skewness by employing a data-driven choice of the distribution of u_i, i.e.:

$$u_i \sim N_{[0,\infty)}(0, \gamma^2) \quad \text{or} \quad u_i \sim N_{[0,a_0|\gamma|]}(a_0|\gamma|, \gamma^2). \tag{15}$$

The latter case has not been explored in the empirical development literature, providing a novel perspective to uncover structural inefficiencies in firm performance in Vietnam. Although we label it as a 'data-driven choice', the sign of γ is not estimated a priori but is determined simultaneously with all other parameters (note that we also allow for $\gamma = 0$). This approach is adopted because any predetermination of residual skewness could be influenced by potential endogeneity. In this context, we examine the possibility of correlated production inputs with composed errors, denoted as $e_i = v_i - u_i$. Endogeneity of inputs can arise due to their correlation with v_i, with u_i, or both. The presence of omitted variables in the production function, such as subsidies or governmental grants that have substantial effects, may lead to correlation with v_i. Furthermore, if producers possess prior knowledge of potential inefficiencies in output generation, they are likely to adjust their inputs accordingly (Haschka and Herwartz 2022). As these adjustments are unobserved by the analyst, they introduce correlation with u_i.

We conduct a comparative analysis of the proposed estimator, which accommodates endogeneity of inputs and considers both correct and wrong skewness of inefficiency, with the alternative estimators considered in the Monte Carlo simulations. These are ML (Hafner et al. 2018), copula (Tran and Tsionas 2015), and GMM (Tran and Tsionas 2013). To set up the IV-based GMM estimator, we

Table 3 Estimation outcomes are presented employing MLE (Hafner et al. 2018), GMM (Tran and Tsionas 2013), copula-based estimation (Tran and Tsionas 2015), and our proposed estimator. Standard errors for copula-based estimators (copula and proposed) are determined through bootstrap procedures with 1,999 replications. Efficiency scores are computed utilising the estimator by Jondrow et al. (1982)

	MLE		GMM		Copula		Proposed	
	Est	SE	Est	SE	Est	SE	Est	SE
const	1.497	.0625	1.241	.0798	1.147	.0802	.9822	.0803
log *wages*	.9229	.0120	.6440	.0204	.6024	.0209	.6833	.0214
log *assets*	.2163	.0099	.2515	.0193	.2904	.0205	.3102	.0211
σ_v	1.082	.0142	.9914	.0295	1.194	.0308	.8993	.0301
γ	.4151	.0312	1.241	.0396	1.290	.0411	-.5809	.0411
$\rho_{e,\log wages}$					.3025	.0403	.2627	.0401
$\rho_{e,\log assets}$					.1790	.0371	.2298	.0366
Mean $\mathbb{E}[u|e]$	.3174		.8747		.9407		.4900	
Mean Efficiency	.7280		.4169		.3903		.6126	

employ one-year lagged assets and one-year lagged wages as instruments, following a similar approach used in previous studies (Haschka and Herwartz 2022). However, it is important to note that such internal instrumentation may suffer from weak instruments and might not be entirely suitable for handling endogeneity.

Table 3 presents the estimation results, which consistently indicate human capital as the primary driver of firm performance in Vietnam across all employed estimators. This is evident from the significantly higher coefficient associated with log wages compared to log assets. The consideration of endogeneity through GMM, Copula, and the proposed estimator reduces the observed difference in coefficients. GMM and Copula estimators may still face remaining endogeneity issues when wrong skewness is present, with GMM encountering additional challenges due to potential weak instrumentation. MLE indicates increasing returns to scale, as reflected in the sum of the elasticities associated with wages and assets being above one. However, for Copula and GMM, this sum significantly falls below one, suggesting decreasing returns to scale. This implies that scaling up is challenging for firms. By contrast, the proposed estimator yields constant returns to scale, which aligns with the dataset dominated by small firms (O'Toole and Newman 2017).

The discrepancy raises concerns about the MLE results being flawed due to endogeneity. Further evidence supporting endogeneity is found in the significant estimates of correlations between production inputs and errors when using Copula and the proposed estimators. In terms of firm efficiency, MLE yields surprisingly high mean efficiency, with an average score of .7280. However, accounting for endogeneity while assuming correct skewness through GMM and Copula estimators decreases mean efficiency scores to .4169 (GMM) and .3903 (Copula). The proposed approach reveals a mean efficiency of .6126. While GMM and Copula estimators exhibit minimal differences among all coefficients, the results undergo significant changes when the proposed estimator is applied. Notably, the proposed

estimator indicates the presence of wrong skewness after accounting for endogeneity, which is challenging for MLE to detect under such conditions.

The insights derived from the proposed estimator contribute evidence supporting the existence of endogenous regressors and wrong skewness. While the former has been previously emphasized in empirical development literature, the latter has not received recognition. To comprehend the economic reasons and market mechanisms behind the occurrence of wrong skewness, it is crucial to delve into the contributing factors. As illustrated in Panel (a) of Fig. 1, the presence of correct skewness (dotted line) indicates that most firms should operate near the efficiency frontier, aligning with competitive market conditions where inefficient firms are likely to exit due to a lack of competitiveness. Conversely, empirical evidence supporting wrong skewness (straight line) implies a growing number of inefficient firms persisting in the market, contradicting the assumption of a competitive market situation. This suggests a lack of incentives for firms to optimize efficiency, attributed to factors such as widespread corruption and the constraints imposed by the communist regime in Vietnam. Despite ongoing reforms, these challenges persist, necessitating policy interventions to incentivize firms to optimize processes and improve efficiency, as market forces alone seem insufficient to generate such incentives.

5 Discussion of the methodology

The proposed approach offers several advantages that make it useful for handling both endogeneity and skewness issues. Firstly, it overcomes the need for instrumental variables, eliminating the challenge of obtaining and validating such instruments. By employing a copula function to directly connect endogenous regressors and errors, the model exhibits parsimony and requires the identification of only one additional parameter for each endogenous regressor - the correlation coefficient depicting regressor-error dependence. This parsimony is further bolstered by the one-parameter inefficiency distribution, which can accommodate both correct and wrong skewness with only one parameter to estimate. The sign of this parameter determines the skewness. Additionally, the likelihood function is readily obtained even under wrong skewness, as the errors follow a CSN distribution for which certain properties can be derived (Flecher et al. 2009). This simplifies the estimation process and enhances the model's computational efficiency.

The proposed approach does have certain limitations, which stem from either the SF specifications or the copula function employed for endogeneity correction. On the one hand, while we attribute wrong skewness to the inefficiency component, other potential sources, such as asymmetry in idiosyncratic noise or dependence between idiosyncratic noise and inefficiency, could also be responsible. Additionally, we restrict our attention to the (truncated) half-normal distribution for inefficiency. While this choice simplifies the analysis by ensuring a CSN distribution after convolution with the normal distribution assumed for idiosyncratic noise, alternative distributions for inefficiency could be considered, such as the negative skew exponential distribution (Hafner et al. 2018). While the choice of appropriate distribution assumptions is a key consideration in any SF model, in the context of wrong

skewness, we would also need to establish whether the density of $e = v - u$, resulting from the convolution of v with u, follows a known parametric distribution.

On the other hand, while the Gaussian copula exhibits considerable flexibility, particularly when dealing with multiple endogenous regressors as in the empirical application, it is rooted in the assumption of linear dependency between its margins. However, prior studies have demonstrated the robustness of the Gaussian copula in capturing various non-Gaussian dependencies (Haschka 2022b; Haschka and Herwartz 2022; Park and Gupta 2012). By virtue, any copula-based endogeneity correction generally disregards the potential sources of endogeneity, be it omitted variables, reverse causality, or simultaneity, as they are employed to address the symptoms of endogeneity. Regarding the SF specification, it is further complicated by the fact of discerning whether endogeneity arises from correlation with the inefficiency term or idiosyncratic noise because we model the joint distribution of endogenous regressors and composite errors. Moreover, the proposed approach only allows for continuous endogenous variables. Since model identification necessitates marginal distributions of endogenous regressors to be different from that of the composite errors (which follow a CSN distribution), empirical applications demand an a priori assessment of the marginal distribution of explanatory variables. For instance, in cases where the model incorporates only fully efficient firms, the endogenous regressors must exhibit a non-normal distribution.

While instrumental variable estimation remains the preferred method for addressing endogeneity when strong and valid instruments are available, weak instrumentation or skewness misspecifications pose empirical challenges to such fully parametric approaches. Nevertheless, we expect that the proposed method offers valuable alternatives for numerous empirical SF models that encounter regressor endogeneity and/or skewness issues.

6 Conclusion

Traditional stochastic frontier (SF) models typically assume that inefficiency follows a half-normal distribution with positive skewness. However, when true inefficiency exhibits negative skewness, efficiency scores are biased toward one, leading to misleading conclusions of high efficiency (Waldman 1982). While recent literature has highlighted the importance of addressing the 'wrong' skewness problem in SF analysis (Curtiss et al. 2021; Choi et al. 2021; Daniel et al. 2019), existing studies have not considered the potential endogeneity of regressors. This paper fills this gap by proposing an instrument-free approach for estimating SF models with endogenous regressors and allowing for a simultaneous choice of inefficiency skewness.

Building upon the work of Park and Gupta (2012), we employ a copula function to directly construct the joint density of endogenous regressors and composite errors, enabling us to capture mutual dependency without the need for instrumental variables. Our model distinguishes between correct and wrong skewness without imposing a priori restrictions on the sign of inefficiency skewness or requiring the identification of additional parameters governing its direction. We evaluate the finite sample performance of the approach through Monte Carlo simulations. The

simulation results demonstrate that the estimator performs well in finite samples, exhibiting desirable properties in terms of bias and mean squared error. These findings are further validated in an empirical application.

The following contributions of this article are also worth mentioning. On the methodological front, we advance the understanding of copula-based endogeneity corrections in scenarios with non-Gaussian outcomes. Joint estimation using copulas still occupies a niche in endogeneity-robust modeling (Papies et al. 2023; Papadopoulos 2021). Our work contributes to bridging this gap by understanding copula-based endogeneity corrections in SF settings.

The existing methodological literature about dealing with skewness issues in SF models predominantly addresses the symptoms but fails to delve into the underlying causes, often attributing skewness issues to weak samples (Almanidis and Sickles 2011; Hafner et al. 2018; Simar and Wilson 2009). Empirically, Papadopoulos and Parmeter (2023) note that only two studies in the past 25 years have considered skewness issues in SF analyses. However, these studies remain silent on the potential economic explanations for skewness. Our study breaks new ground by attributing skewness to the inefficiency component of the SF model, providing an economic explanation for this phenomenon.

Regarding the empirical application, this study is the first to explicitly address and account for skewness issues when applying SF models in a development economics context. Previous studies on firm growth in Vietnam using SF analysis, such as Haschka et al. (2023), have overlooked potential skewness issues in their findings. By incorporating an economic explanation for wrong skewness and proposing an endogeneity-robust methodology, we offer a novel perspective on growth dynamics in Vietnam and shed light on underlying competition levels.

References

Almanidis P, Sickles RC (2011) The skewness issue in stochastic frontiers models: fact or fiction? In: van Keilegom I, Wilson PW (eds) Exploring research frontiers in contemporary statistics and econometrics. Springer, Berlin, pp 201–227

Amsler C, Prokhorov A, Schmidt P (2016) Endogeneity in stochastic frontier models. J Econom 190(2):280–288

Badunenko O, Henderson DJ (2023) Production analysis with asymmetric noise. J Product Anal 61:1–18

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir Econ 20(2):325–332

Becker J-M, Proksch D, Ringle CM (2021) Revisiting gaussian copulas to handle endogenous regressors. J Acad Market Sci (forthcoming)

Bonanno G, Domma F (2022) Analytical derivations of new specifications for stochastic frontiers with applications. Mathematics 10(20):3876

Bonanno G, De Giovanni D, Domma F (2017) The 'wrong skewness' problem: a re-specification of stochastic frontiers. J Product Anal 47(1):49–64

Breitung J, Mayer A, Wied D (2023) Asymptotic properties of endogeneity corrections using nonlinear transformations. Econom J, utae002

Chen Y-Y, Schmidt P, Wang H-J (2014) Consistent estimation of the fixed effects stochastic frontier model. J Econom 181(2):65–76

Choi K, Kang HJ, Kim C (2021) Evaluating the efficiency of Korean festival tourism and its determinants on efficiency change: parametric and non-parametric approaches. Tourism Manag 86:104348

Curtiss J, Jelínek L, Medonos T, Hruška M, Hüttel S (2021) Investors' impact on Czech farmland prices: a microstructural analysis. Eur Rev Agric Econ 48(1):97–157

Daniel BC, Hafner CM, Simar L, Manner H (2019) Asymmetries in business cycles and the role of oil prices. Macroecon Dyn 23(4):1622–1648

Flecher C, Naveau P, Allard D (2009) Estimating the closed skew-normal distribution parameters using weighted moments. Stat Probab Lett 79(19):1977–1984

Genest C, Ghoudi K, Rivest L-P (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82(3):543–552

Green A, Mayes D (1991) Technical inefficiency in manufacturing industries. Econ J 101(406):523–538

Hafner CM, Manner H, Simar L (2018) The wrong skewness problem in stochastic frontier models: a new approach. Econom Rev 37(4):380–400

Haschka RE (2021) Exploiting between-regressor correlation to robustify copula correction models for handling endogeneity. SSRN Working Paper. https://ssrn.com/abstract=4222808

Haschka RE (2022a) Bayesian inference for joint estimation models using copulas to handle endogenous regressors. SSRN Working Paper. https://ssrn.com/abstract=4235194

Haschka RE (2023) Endogeneity-robust estimation of nonlinear regression models using copulas: a Bayesian approach with an application to demand modelling. SSRN Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4451591

Haschka RE, Wied D (2022) Estimating fixed effects stochastic frontier panel models under 'wrong' skewness with an application to health care efficiency in Germany. SSRN Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4079660

Haschka RE (2022) Handling endogenous regressors using copulas: A generalisation to linear panel models with fixed effects and correlated regressors. J Market Res 59(4):860–881

Haschka RE, Herwartz H (2020) Innovation efficiency in European high-tech industries: evidence from a Bayesian stochastic frontier approach. Res Policy 49:104054

Haschka RE, Herwartz H (2022) Endogeneity in pharmaceutical knowledge generation: an instrument-free copula approach for Poisson frontier models. J Econ Manag Strategy 31(4):942–960

Haschka RE, Schley K, Herwartz H (2020) Provision of health care services and regional diversity in Germany: insights from a Bayesian health frontier analysis with spatial dependencies. Eur J Health Econ 21:55–71

Haschka RE, Herwartz H, Struthmann P, Tran VT, Walle YM (2021) The joint effects of financial development and the business environment on firm growth: evidence from Vietnam. J Comp Econ 50(2):486–506

Haschka RE, Herwartz H, Silva Coelho C, Walle YM (2023) The impact of local financial development and corruption control on firm efficiency in Vietnam: evidence from a geoadditive stochastic frontier analysis. J Product Anal 60(2):203–226

Horrace WC, Parmeter CF, Wright IA (2023) On asymmetry and quantile estimation of the stochastic frontier model. J Product Anal 61(1):19–36

Joe H, Xu JJ (1996) The estimation method of inference functions for margins for multivariate models. Technical Report: The University of British Columbia, Canada

Johnson NL, Kotz S, Balakrishnan N (1995) Continuous univariate distributions, vol 2. Wiley, New York

Jondrow J, Lovell CK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econom 19(2–3):233–238

Kumbhakar SC, Lovell CK (2003) Stochastic frontier analysis. Cambridge University Press, Cambridge

Kutlu L (2010) Battese–Coelli estimator with endogenous regressors. Econ Lett 109(2):79–81

Li Q (1996) Estimating a stochastic production frontier when the adjusted error is symmetric. Econ Lett 52(3):221–228

O'Toole C, Newman C (2017) Investment financing and financial development: Vidence from Vietnam. Rev Finance 21(4):1639–1674

Papadopoulos A (2021) Measuring the effect of management on production: a two-tier stochastic frontier approach. Empir Econ 60(6):3011–3041

Papadopoulos A (2022) Accounting for endogeneity in regression models using Copulas: a step-by-step guide for empirical studies. J Econom Methods 11(1):127–154

Papadopoulos A, Parmeter CF (2023) The wrong skewness problem in stochastic frontier analysis: a review. J Product Anal. https://doi.org/10.1007/s11123-023-00708-w

Papies D, Ebbes P, Feit EM (2023) Endogeneity and causal inference in marketing. In: Winder RS, Neslin SA (eds) The history of marketing science. World Scientific Publishing Co. Pte. Ltd., Singapore, pp 253–300

Park S, Gupta S (2012) Handling endogenous regressors by joint estimation using Copulas. Market Sci 31(4):567–586

Prokhorov A, Schmidt P (2009) Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. J Econom 153(1):93–104

Prokhorov A, Tran KC, Tsionas MG (2020) Estimation of semi-and nonparametric stochastic frontier models with endogenous regressors. Empir Econ 60:3043–3068

Simar L, Wilson PW (2009) Estimation and inference in cross-sectional, stochastic frontier models. Econom Rev 29(1):62–98

Smith MD (2008) Stochastic frontier models with dependent error components. Econom J 11(1):172–192

Son TVH, Coelli T, Fleming E (1993) Analysis of the technical efficiency of state rubber farms in Vietnam. Agric Econ 9(3):183–201

Tran KC, Tsionas MG (2021) Efficient semiparametric copula estimation of regression models with endogeneity. Econom Rev (forthcoming)

Tran KC, Tsionas EG (2013) GMM estimation of stochastic frontier models with endogenous regressors. Econ Lett 118:233–236

Tran KC, Tsionas EG (2015) Endogeneity in stochastic frontier models: copula approach without external instruments. Econ Lett 133:85–88

Tsionas MG (2017) When, where, and how of efficiency estimation: improved procedures for stochastic frontier modeling. J Am Stat Assoc 112(519):948–965

Waldman DM (1982) A stationary point for the stochastic frontier likelihood. J Econom 18(2):275–279