A SEMIPARAMETRIC CONSTANT ELASTICITY OF SUBSTITUTION STOCHASTIC FRONTIER MODEL FOR PANEL DATA[†]

TAINING WANG

International School of Economics and Management
Capital University of Economics and Business
Beijing, 100070, China
Email: taining.wang@cueb.edu.cn

DANIEL J. HENDERSON

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

This version: December 23, 2022

Abstract: We propose a semiparametric stochastic frontier model for panel data with several flexible features. First, we start with a constant elasticity of substitition (CES) production frontier, and estimate it without log-transformation to avoid induced non-negligible estimation bias. Second, we improve flexibility via semiparameterization, where the technology is an unknown function of a set of environment variables. The technology function further accounts for latent heterogeneity across individual units, which can be freely correlated with inputs, environment variables, and/or inefficiency determinants. We incorporate a single-index structure in the technology function to circumvent the curse of dimensionality. Third, we eschew distributional assumptions on both stochastic noise and inefficiency for model identification. We only assume the conditional mean of the inefficiency depends on related determinants, which permits a wide range of choice, via a positive parametric function. As a result, technical efficiency in our model is constructed without relying on an assumed distribution on composite error. Our model provides flexible structures on both the production frontier and inefficiency, thereby alleviating the risk of model misspecification in production and efficiency analysis. Our estimator involves a series based nonlinear least squares estimation for the unknown parameters and a kernel based local estimation for the technology function. We showcase the promising finite-sample performance via simulations, and apply the model to investigate productive efficiency among OECD countries from 1970-2019.

Keywords: panel data; semiparametric model; CES; stochastic frontier; profile nonlinear least-square; local nonlinear kernel

JEL Classifications: C14, C23

[†]The authors would like to thank the co-editor, Chris Parmeter, and two anonymous referees for helpful comments and suggestions. An R package which can be used to apply the methods developed in this paper is available from the authors' websites and discussed in the Appendix B.

1 Introduction

The seminal works by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) evoked a myriad of studies to estimate production functions and inefficiency via stochastic frontier models (SF).[1] The basic structure of the SF specifies a production (frontier) function $Y = m(X)$ for output Y and input vector X in the presence of a composite error $e = v - u$. Such structure is practically appealing as $m(X)$ can be specified to perform inference of economic interest and e distinguishes unobserved output variation due to stochastic noise v (e.g., production shocks, measurement error) from that due to productive inefficiency u (e.g., managerial inefficiency). The function $m(\cdot)$ is often assumed to be Cobb-Douglas (CD) or constant elasticity of substitution function (CES). Unlike a regression model with zero mean noise, the composite error has a non-zero mean due to the one-sided nature of $u > 0$ (i.e., the shortfall of the output from its frontier). Originally, parametric distributional assumptions on v (e.g., normal) and u (e.g., half-normal, exponential, truncated normal, gamma) were required for identification.

Parametric assumptions on both $m(X)$ and e are often made via conjecture, for which violations result in inconsistent estimation and potentially invalid policy implications (Kumbhakar et al., 2020). Effort has been made to relax these restrictive assumptions. To relax the parameteric assumptions on the distribution of e, panel data methods are often employed. For example, Schmidt and Sickles (1984) consider time-invariant inefficiency, which can be bundled and eliminated with additive fixed effects. Cornwell et al. (1990) model inefficiency as a deterministic function of a time trend. Kumbhakar (1990), Battese and Coelli (1992), Lee and Schmidt (1993), and Kumbhakar and Wang (2005a) specify a time-invariant inefficiency, scaled by a deterministic function of time with different functional forms. Alternatively, Horrace and Parmeter (2011) estimate the unknown distribution of the composite error via deconvolution, and Tran and Tsionas (2009), Parmeter et al. (2016) and Zhou et al. (2020) adopt a partially linear SF to estimate the mean function of inefficiency with distribution free composite error. However, the aforementioned studies relax the parametric distribution assumption at the cost of maintaining a parametric frontier structure. A correctly specified frontier model can be paramount to provide reliable insights on many important features of the production function, such as returns to scale.

Kneip and Simar (1996) and Fan et al. (1996) were the first to consider a fully nonparametric frontier model $m(X)$. Martins-Filho and Yao (2015) investigate the asymptotic properties of the estimator of Fan et al. (1996), and propose a refined profile-likelihood based estimator for the nonparametric frontier. Kumbhakar et al. (2007) model all parameters in the frontier and composite error distribution nonparametrically with a kernel weighted local log-likelihood approach. Park et al. (2015) further generalize their results to incorporate categorical data in $m(X)$. Kneip et al. (2015) propose a fully nonparametric frontier model when v is log-normally distributed with un-

[1] For comprehensive reviews of SF models, see Kumbhakar and Lovell (2000), Parmeter and Kumbhakar (2014), and Kumbhakar et al. (2020). For the illustration of SF empirical studies, see Kumbhakar et al. (2015).

known variance. In practice, the fully nonparametric frontier can be imprecisely estimated due to the well-known *curse of dimensionality* problem. To circumvent this curse, recent studies employ semiparametric techniques: additive (Ferrara and Vidoli, 2017) or varying coefficient (Sun and Kumbhakar, 2013; Yao et al., 2018a) structures. Even with these improvements, these non/semiparametric models maintain parametric distribution assumptions on e, where violations can lead to model identification problem (Parmeter et al., 2016; Kumbhakar et al., 2020).

The restrictive parametric assumptions on e or $m(X)$ have been relaxed on one side at the expense of being maintained on the other side. In this paper, we allow for some structure, but we relax some assumptions on both sides by proposing a semiparametric panel frontier model which unifies several attractive features. First, we do not impose distributional assumptions on v and u. We allow u to be influenced by observed determinants W through its mean function, whose structure is generally unknown.[2] As the number of inefficiency determinants rises (see, for example, Iyer et al. (2008)), nonparametric estimation on the unknown mean function suffers from the curse of dimensionality. To improve the model's applicability, we assume the mean function of W to be positive and known up to a finite dimensional parameter vector. As emphasized by Parmeter and Kumbhakar (2014), a correct modeling on the dependence of u on W is important to avoid biasing the estimates of both frontier and technical efficiency. We further provide discussion on technical efficiency measures suitable in our model without distributional assumptions. Different from existing non/semiparametric frontier models that often place restrictions on the choice of W (Ferrara and Vidoli, 2017; Sun and Kumbhakar, 2013; Yao et al., 2018a), we allow W to be chosen in a flexible fashion, which can include all, some, or none of variables in the frontier function.

Second, we begin by specifying the frontier as a semiparametric extension of CES. In a production function with two inputs K (physical capital) and L (labor), the classical (parametric) CES with constant returns to scale specifies $A\left[\pi K^{-\rho}+(1-\pi)L^{-\rho}\right]^{-\frac{1}{\rho}}$, where $A>0$ is the technology parameter, π and $(1-\pi)$ are the shares of K and L, respectively, and ρ determines the elasticity of substitution $\sigma=\frac{1}{1+\rho}$ between K and L. The flexible structure of CES receives considerable attention in the literature because, as a generalization of CD, it allows σ to differ from unity and allows non-constant factor shares (Golin, 2002). Here, we further improve the flexibility of CES by allowing A to be an unknown function of environment variables Z (i.e., $A=A(Z)$), which are not traditional production inputs but generate facilitating impacts on Y, such as computerization (Yao et al., 2018a), human capital (Yao et al., 2018b), or firm's debt (Wang et al., 2020). To avoid the curse of dimensionality, we adopt a single-index structure as $A(Z^{\top}\gamma_z)$, where γ_z is a vector of coefficients (Ma and Song, 2015).

This general form has at least three advantages over existing models. First, parametric or semiparametric models typically involve log-transformations. Assuming e is multiplicative, the

[2] Tran and Tsionas (2009), Parmeter et al. (2016) and Zhou et al. (2020) estimate the mean function nonparametrically, which is appealing.

parametric CES is estimated by first logging both sides of the SF and then approximated by Taylor expansions (i.e., Kmenta approximation). Similarly, the semiparametric varying coefficient frontier can be viewed as a generalization of a log-transformed CD (Wang et al., 2020). As suggested by Sun et al. (2008), log transformations on $m(\cdot)$ can lead to inconsistent estimation, since the expectation of the log error term is almost always a function of X (or (X, Z, W) in our case) (Santos Silva and Tenreyro, 2006). We avoid the conventional practice of log transformations when estimating our semiparametric CES, and follow Wang and Henderson (2022) to estimate the model directly in levels. Second, the environment variables Z can be high-dimensional, thereby facilitating investigation on the joint effects of multiple Z in the technology function without inflating estimation inaccuracy. Third, different from non/semiparametric models, our frontier allows direct estimation of some parameters of interests with important economic implications. For instance, estimating the elasticity of substitution (σ) is straightforward in our model, which can potentially explain the declining share of labor income in recent decades (Karabarbounis and Neiman, 2013). Such parameters are difficult to identify in semi/nonparametric models (Henderson, 2009). Our frontier specifies the technology function flexibility while adopting parsimonious structure on modeling certain features of the production function.

Finally, we control for unobserved and time invariant individual heterogeneities, which is often absent in non/semiparametric models. It is important to incorporate fixed effects in panel SF models as it provides an efficient way to mitigate potential endogeneity. We allow the fixed effects to be arbitrarily correlated with covariates (X, Z, W), and estimate them via dummy variables in the index of the technology function $A(\cdot)$, analogous to a logged CES model where the fixed effects appear as intercept shifts.

We propose a two-step estimation procedure to estimate our semiparametric panel SF. In the first step, we employ a profile nonlinear least-square estimator to estimate the global parameters in the frontier and inefficiency functions, where the unknown function $A(\cdot)$ is approximated by a series estimator. In the second step, we employ a local nonlinear kernel estimator for $A(\cdot)$, using all parameter estimates from the previous step. Our simulation study shows that the proposed estimator works fairly well in finite samples.[3] Using a panel dataset of 20 OECD countries from 1970-2019, our model is demonstrated to be more suitable than conventional parametric SF models to uncover important features of production frontier and inefficiency.

The remainder of this chapter proceeds as follows: we present the proposed estimator and identification conditions in Section 2, illustrate the finite sample performance through simulations studies in Section 3, demonstrate an empirical application in Section 4, and conclude in Section 5. The appendicies include additional simulations (detailing the incidental parameters problem) as well as a description of the R package we created for this estimator.

[3]Similar to the findings in Wang and Henderson (2022), we will suggest, in Section 2.3, large n and T in practice in order to avoid the incidental parameters problem.

2 Semiparametric CES Stochastic Frontier Model

2.1 Model

We consider the following semiparametric CES stochastic frontier model (SF-CES-SP):

$$Y_{it} = A(Z_{it}^{\top}\gamma_z + \alpha_i)\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho}\right)^{-\frac{\nu}{\rho}} + v_{it} - u_{it},\ i = 1, ..., n,\ t = 1, ..., T, \quad (1)$$

where n and T represent the total individual and time units, respectively. $Y_{it} \in \Re_+$ is the univariate production output and $X_{it} \in \Re_+^{d_x}$ are traditional production inputs. A linear combination of environment variables $Z_{it} \in \Re^{d_z}$ influences the technology $A(\cdot)$ by a scale effect, where $A(\cdot)$ is an unknown smooth function. We assume the fixed effects α_i appear in the SF-CES through $A(\cdot)$ such that $E(\alpha_i|X_{it}, Z_{it}, Q_{it}) \neq 0$ (Baltagi and Li, 2002), where $Q_{it} \in \Re^{d_q}$ are pure inefficiency determinants, such as export intensity (Iyer et al., 2008), which do not show up in the frontier. We incorporate α_i via dummy variables, and for identification purposes, we include all but the first dummy so that a total of $d_z + (n-1)$ variables are included in $A(\cdot)$. The parameter ρ $(-1 < \rho \neq 0)$ determines the elasticity of substitution $\sigma = \frac{1}{1+\rho}$ (assumed the same) between two production inputs X_j and X_s for $j \neq s$. $\nu > 0$ is the returns to scale and $0 \leq \pi_j \leq 1$ is the distribution parameter that explains the relative input share $(\sum_{j=1}^{p} \pi_j = 1)$.[4]

In the composite error, the one-sided error $u_{it} > 0$ is the time-variant inefficiency variable, measuring the difference between the maximum output produced at the frontier and the actual output produced.[5] We do not impose a particular distribution on u_{it}, but only assume the existence of its positive conditional mean function $E(u_{it}|\alpha_i, X_{it}, Z_{it}, Q_{it}) \equiv \mu(W_{it}; \theta) > 0$ known up to the finite dimensional parameter $\theta \in \Re^{d_w}$, with the determinants W_{it} being a subset of X_{it}, Z_{it}, and Q_{it}, so W_{it} is of dimension $d_w \leq d_x + d_z + d_q$. Finally, the two-sided error v_{it} represents random shocks in production process, which satisfies $E(v_{it}|\alpha_i, X_{it}, Z_{it}, Q_{it}) = 0$ for all i and t.

Our model improves existing SF models in the literature by relaxing restrictive assumptions on both the frontier and composite error. The semiparametric CES generalizes traditional parametric CD and CES by specifying the technology to be dependent on Z_{it} in an unknown fashion. Our frontier also permits large dimension of Z_{it} without affecting estimation accuracy, and allows the presence of interactive effects among inputs as opposed to other semiparametric varying coefficient frontier models (Ferrara and Vidoli, 2017; Sun and Kumbhakar, 2013; Yao et al., 2018a,b; Wang et al., 2020). More importantly, the CES frontier is estimated in levels to avoid the bias due to log-transformations which exists in most parametric/semiparametric SF models. Our composite error is free of distribution misspecification as we only assume that the conditional mean of u_{it} exists.

[4] Note that π_j in (1) is not the output elasticity of the j^{th} input as is the case in CD.

[5] We mention that our model specification does not account for time invariant, or persistent, inefficiency such that $u_{it} = c_i^* + u_{it}^*$ with $c_i^*, u_{it}^* > 0$. Such decomposition receives increasing attention in parametric SF panel models. See Kumbhakar et al. (2020) for more details.

Further, the choice of determinants W_{it} in $\mu(\cdot;\theta)$ is flexible as our production inefficiency may emanate from X_{it}, Z_{it}, and/or Q_{it} (Parmeter and Kumbhakar, 2014). This feature contrasts, and may therefore ameliorate, the limited choices of W_{it} in existing semiparametric models ($W_{it} = \emptyset$ in additive (Ferrara and Vidoli, 2017), $W_{it} = Z_{it}$ in varying coefficient (Yao et al., 2018a), and $W_{it} = Q_{it}$ in partially linear (Tran and Tsionas, 2009; Parmeter et al., 2016; Zhou et al., 2020) models).

Our SF-CES-SP in (1) is clearly not a regression model because the conditional mean of $v_{it} - u_{it}$ is not zero. It can be readily transformed into a regression model by de-meaning u_{it} in the composite error such that $\epsilon_{it} = v_{it} - (u_{it} - \mu(W_{it};\theta))$. Thus, a regression-type SF-CES-SP is given as

$$Y_{it} = A(Z_{it}^{\top}\gamma_z + \alpha_i)\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho}\right)^{-\frac{\nu}{\rho}} - \mu(W_{it};\theta) + \epsilon_{it}, \tag{2}$$

where $E(\epsilon_{it}|\alpha_i, X_{it}, Z_{it}, Q_{it}) = 0$ is now satisfied by construction. Below, we discuss the identification conditions for parameters from (2) in Section 2.2, propose a two-step model estimator in Section 2.3, and list basic assumptions for establishing the estimator's consistency in Section 2.4.

2.2 Identification

We first discuss identification of parameters in (2). Let $\alpha_{-1} = (\alpha_2, ..., \alpha_n)^{\top}$ be a vector that collects all but the first fixed effect. Define $\gamma = (\gamma_z^{\top}, \alpha_{-1}^{\top})^{\top}$ and $\omega = (\pi_1, ..., \pi_{d_x}, \rho, \nu, \theta^{\top})^{\top}$, we assume that the global parameters $\eta = (\gamma^{\top}, \omega^{\top})^{\top} \in \Theta$, where $\Theta = \{\Gamma, \Omega\}$ is a collection of compact parameter spaces in which $\gamma \in \Gamma$ and $\omega \in \Omega$.

To identify γ in $A(\cdot)$, we follow Ma and Song (2015) and let γ belong to a compact parameter space $\Gamma = \{\gamma : ||\gamma|| = 1, \gamma_1 > 0, \gamma_l \in \Re,\ l = 2, ..., d_z + n - 1\}$, where $||m||$ denotes the L_2 norm of any vector m. The condition $||\gamma|| = 1$ ensures $A(\cdot)$ is identifiable by restricting the scale of γ. As a result, the true values of γ locate at the boundary points of a unit sphere, so the derivative of $A(\cdot)$ cannot be evaluated at γ. Following Zhu and Xue (2006), we let all but the first element in γ be defined on the interior points of the unit sphere by assuming $||\gamma_{-1}|| < 1$ with $\gamma_{-1} = [\gamma_{z_2}, ..., \gamma_{z_{d_z}}, \alpha_{-1}^{\top}]^{\top} \in \Re^{d_z+n-2}$, so $\gamma_{z_1} = (1 - ||\gamma_{-1}||^2)^{1/2} > 0$ is the result given in Γ.[6] We consider a "leave-one-out" parameter space for γ as

$$\Gamma_{-1} = \{\gamma_{-1} :\ ||\gamma_{-1}|| < 1, \gamma_{l,-1} \in \Re, l = 2, ..., d_z + n - 1\}$$

from which γ can be uniquely defined.

To identify ω, we assume ω belongs to a compact parameter space $\Omega = \{\rho, \nu, \theta, \pi = (\pi_j : 1 \leq j \leq d_x) : -1 < \rho \neq 0, \nu > 0, \theta \in \Re^{d_w}, 0 \leq \pi_j \leq 1, \sum_{j=1}^{d_x} \pi_j = 1\}$. Given that $\sum_{j=1}^{d_x} \pi_j = 1$ implies

[6] In practice, one can set any element in γ to be positive for $j = 1, 2, ..., d_z + n - 1$, or to be negative by setting $\gamma_j = -(1 - ||\gamma_{-j}||^2)^{1/2} < 0$.

$\pi_1 = 1 - \sum_{j=2}^{d_x} \pi_l$, we consider a "leave-one-out" parameter space for ω as[7]

$$\Omega_{-1} = \{\rho, \nu, \theta, \pi_{-1} = (\pi_l : 2 \leq l \leq d_x) : -1 < \rho \neq 0, \nu > 0, \theta \in \Re^{d_w}, 0 \leq \pi_l \leq 1, 0 \leq \sum_{l=2}^{d_x} \pi_l \leq 1\}.$$

The conditions above allow us to identify $\eta_{-1} = (\gamma_{-1}^\top, \omega_{-1}^\top)^\top$ from the resulting "leave-one-out" parameter space $\Theta_{-1} = \{\Gamma_{-1}, \Omega_{-1}\}$.

We must point out that in a rare case, γ_{-1} cannot be differentiated from Θ_{-1}. Assume that $\alpha_i = 0$ for all i and that Z_{it} are deemed to generate the same effects on both the technology and inefficiency (i.e., $W_{it} = Z_{it}$ and $\gamma_z = \theta$). Then if $A(\cdot)$ takes the same functional form as $\mu(\cdot)$, we have $A(Z_{it}^\top \gamma_z) = \mu(W_{it}; \theta)$. The frontier in (2) reduces to $A(Z_{it}^\top \gamma_z) \left[\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho} \right)^{-\frac{\nu}{\rho}} - 1 \right]$, where γ_{-1} cannot be identified if $\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho} \right)^{-\frac{\nu}{\rho}} = 1$ for some particular values of X_j and (ρ, π_j, ν). We are less concerned about this identification problem because the probability of having zero fixed effects, identical effects of Z_{it} in frontier and inefficiency, and equivalent structures of $A(\cdot)$ and $\mu(\cdot)$, is essentially zero. Even such an event occurs, we can let $\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho} \right)^{-\frac{\nu}{\rho}} \neq 1$ by requiring $X_j > 1$ for all j, which almost always holds in applied studies.

2.3 Estimation

We employ a two-step semiparametric estimator for our model in (2). In the first step, we focus on estimating $\eta = (\gamma^\top, \omega^\top)$ using series approximation. If η were known, the only unknown part in (2) is the smooth technology function $A(\cdot)$, which can be approximated via a series estimator. To avoid clustering notations below, we denote $\mathbf{Z} \in \Re^{d_z + n - 1}$ as a vector including environment variables Z plus $n-1$ dummy variables for fixed effects. Let $\mathcal{Z}_j = [a_j, b_j] \in \Re$ be a compact subset of $\mathbf{Z}_j$ for some real constants a_j and b_j, with $j = 1, ..., d_z + n - 1$. For $\mathbf{z} = [\mathbf{z}_1, ..., \mathbf{z}_{d_z+n-1}]^\top$ with $\mathbf{z}_j \in \mathcal{Z}_j$, let $v(\gamma) = \mathbf{z}^\top \gamma$. We define $\phi(v(\gamma)) = [\phi_1(v(\gamma)), ..., \phi_L(v(\gamma))]^\top$ as a $L \times 1$ vector of basis functions evaluated at $v(\gamma)$, where L is a positive integer satisfying $L \to \infty$ as $nT \to \infty$. Then $A(v(\gamma)) \approx \phi(v(\gamma))^\top \lambda(\eta)$, where $\lambda(\eta) = [\lambda_1(\eta), ..., \lambda_L(\eta)]^\top \in \Re^L$ are unknown series coefficients. Here, we highlight the dependence of $\lambda(\cdot)$ with η because $A(\cdot)$ is not additively separable with other parametric parts in (2). Let $N = nT$, the estimator of $\lambda(\eta)$ solves the least-square problem

[7] We note that the condition $\sum_{j=1}^{\pi_{d_x}} \pi_j = 1$ is not necessary for ω in (1) to be identified. Rather, the condition is commonly made in the literature to provide economic interpretation on the distribution parameters (Masanjala and Papageorgiou, 2004).

as $\widehat{\lambda}(\eta) = \underset{\{\lambda(\eta)\in\Re^L\}}{\text{argmin}}\ \mathcal{L}_N(\lambda(\eta),\eta^\top)$, where

$$\mathcal{L}_N(\lambda(\eta),\eta^\top) = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[Y_{it} - \phi(v_{it}(\gamma))^\top\lambda(\eta)\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x}X_{d_xit}^{-\rho}\right)^{-\frac{\nu}{\rho}} + \mu(W_{it};\theta)\right]^2, \quad (3)$$

with $\phi(v_{it}(\gamma)) = \phi(\mathbf{Z}_{it}^\top\gamma)$. In practice, the estimator $\widehat{\lambda}(\eta)$ is infeasible because η is unknown. As mentioned above, we cannot estimate η directly through its first order condition in (3) due to the location of γ and the condition $\pi_1+\cdots+\pi_{d_x}=1$. Given the construction of $\widehat{\lambda}(\eta)$, we estimate $\eta_{-1} = (\gamma_{-1},\omega_{-1})$ through a profile nonlinear least-square estimator (PNLS) $\widehat{\eta}_{-1} = \underset{\{\eta_{-1}\in\Theta_{-1}\}}{\text{argmin}}\ \mathcal{L}_N(\eta_{-1})$, where

$$\mathcal{L}_N(\eta_{-1}) = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[Y_{it} - \phi(v_{it}(\gamma))^\top\widehat{\lambda}(\eta)\left(\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x}X_{d_xit}^{-\rho}\right)^{-\frac{\nu}{\rho}} + \mu(W_{it};\theta)\right]^2. \quad (4)$$

Define the short-hand notations $\widehat{A}(\mathbf{Z}_{it}^\top\gamma) = \phi(v_{it}(\gamma))^\top\widehat{\lambda}(\eta)$ and $M(X_{it};\rho,\pi) = \sum_{j=1}^{d_x}\pi_j X_{jit}^{-\rho}$, the estimator in (4) is equivalent to solving the first-order conditions $\frac{\partial\mathcal{L}_N(\eta_{-1})}{\partial\eta_{-1}} = 0$, where

$$\frac{\partial\mathcal{L}_N(\eta_{-1})}{\partial\eta_{-1}} = -\sum_{i=1}^{n}\sum_{t=1}^{T}\left[Y_{it} - \widehat{A}(\mathbf{Z}_{it}^\top\gamma)M(X_{it};\rho,\pi)^{-\frac{\nu}{\rho}} + \mu(W_{it};\theta)\right] \times \begin{bmatrix} M(X_{it};\rho,\pi)^{-\frac{\nu}{\rho}}\left(J^\top\mathbf{Z}_{it}\widehat{A}'(\mathbf{Z}_{it}^\top\gamma) + \frac{\partial}{\partial\gamma_{-1}}\widehat{A}(\mathbf{Z}_{it}^\top\gamma)\right) \\ \widehat{A}(\mathbf{Z}_{it}^\top\gamma)M(X_{it};\rho,\pi)^{-\frac{\nu+\rho}{\rho}}\nu(-\rho X_{2it}^{\rho})^{-1} \\ \vdots \\ \widehat{A}(\mathbf{Z}_{it}^\top\gamma)M(X_{it};\rho,\pi)^{-\frac{\nu+\rho}{\rho}}\nu(-\rho X_{d_xit}^{\rho})^{-1} \\ \widehat{A}(\mathbf{Z}_{it}^\top\gamma)M(X_{it};\rho,\pi)^{-\frac{\nu}{\rho}}\frac{\partial}{\partial\rho}\left(\frac{\nu}{\rho}\ln\left(M(X_{it};\rho,\pi)\right)\right) \\ \widehat{A}(\mathbf{Z}_{it}^\top\gamma)M(X_{it};\rho,\pi)^{-\frac{\nu+\rho}{\rho}}\frac{1}{\rho}\ln\left(M(X_{it};\rho,\pi)\right) \\ -\frac{\partial}{\partial\theta}\mu(W_{it};\theta) \end{bmatrix},$$

where $\widehat{A}'(\mathbf{Z}_{it}^\top\gamma)$ is the series estimator for the first order derivatives (de Boor, 1978), and

$$J \equiv \frac{\partial\gamma}{\partial\gamma_{-1}} = \begin{pmatrix} -(1-||\gamma_{-1}||^2)^{-1/2}\gamma_{-1}^\top \\ I_{d_z+n-2} \end{pmatrix}$$

is a $(d_z+n-1)\times(d_z+n-2)$ Jacobian matrix, with I_v a $v\times v$ identity matrix for a positive integer v. We obtain $\widehat{\gamma}$ via updating $\widehat{\gamma} = [(1-||\widehat{\gamma}_{-1}||^2)^{1/2},\widehat{\gamma}_{z_2},...,\widehat{\gamma}_{z_{d_z}},\widehat{\alpha}_{-1}^\top]^\top$.

In the second step, we estimate $A(\cdot)$ by a local nonlinear kernel estimator (LNK). Using the estimates $\widehat{\eta}$ from the first step, let $a_0 = A(v(\widehat{\eta}))$ and $a_1 = A'(v(\widehat{\eta}))$ be constants in which $A(\cdot)$ and

the its derivative $A'(\cdot)$ are evaluated at $\mathbf{Z}=\mathbf{z}$ with $\widehat{\gamma}$. The LNK $\widehat{a}=(\widehat{a}_0,\widehat{a}_1)^\top \equiv \widehat{a}(\widehat{\eta})$ is obtained from

$$\widehat{\alpha}(\widehat{\eta})^\top = \underset{\{a(\widehat{\eta})^\top\}}{argmin} \sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}-\left[\alpha_0+\alpha_1(\mathbf{Z}_{it}-\mathbf{z})^\top\widehat{\gamma}\right]\left(\sum_{j=1}^{d_x}\widehat{\pi}_j X_{jit}^{-\widehat{\rho}}\right)^{-\frac{\widehat{\nu}}{\widehat{\rho}}}+\mu(W_{it};\widehat{\theta})\right)^2 k\left(\frac{(\mathbf{Z}_{it}-\mathbf{z})^\top\widehat{\gamma}}{h}\right). \tag{5}$$

where $k(\cdot)$ is a univariate, symmetric kernel function with a bandwidth h such that $h\to 0$ as $nT\to\infty$ (Henderson and Parmeter, 2015). The LNK is essentially a local linear estimator performed on nonlinear functions of $v_{it}(\widehat{\eta})$ defined similarly in Wang and Henderson (2022).

Remark 1. The series in our estimation procedure serves only as a pilot estimator for η. The series estimator facilitates η estimation by easily imposing the structure of semiparametric CES frontier in (1). This is essentially owing to is ability to *globally* linearize $A(\cdot)$ by a number of basis functions, which grows towards infinity as $nT\to\infty$ to achieve estimation consistency. The nature of series approximation also makes it computationally efficient for the global parameters η to be estimated. It would be theoretically feasible to replace the series estimator with a modified LNK in (5) to estimate η; however, doing so becomes empirically infeasible as $nT\to\infty$, because the kernel based estimator needs to *locally* linearize $A(\cdot)$ via Taylor expansion at each sample point. Together with the nonlinear (parametric) structure in the frontier, the LNK based estimator for η would significantly increase the computational burden that makes its applicability prohibitive with large nT. Further, once η is consistently estimated at a parametric rate, LNK is employed to approximate $A(\cdot)$ and $A'(\cdot)$ because the asymptotic distribution of the kernel estimator for both the function and derivatives are more transparent than its series counterpart. Using LNK in the second step also helps for testing hypotheses regarding the technology function.

Remark 2. When including fixed-effects (α_i), our estimator suffers from the incidental parameters problem (i.e., inconsistent when n grows toward infinity while T remains fixed). In a parametric nonlinear panel model with individual fixed-effects (e.g., a panel logit model), the parameter estimator is known to have a leading bias term of order $1/T$ (Arellano and Hahn, 2007). We expect such a bias with a similar order to appear in our proposed estimator for $A(\cdot)$. We show in Appendix A that the incidental parameters problem indeed arises with large n and fixed T, which leads to inconsistent estimation on both parameters and the nonparametric function. To avoid the incidental parameters problem, we follow the insights of Arellano and Hahn (2007) and Fernández-Val and Vidoli (2016) to consider $n,T\to\infty$ and a bounded ratio of n/T to achieve consistent estimation.

Remark 3. Technical efficiency (TE) has long been an important and informative indicator in efficiency analysis. In conventional SF models with log output, a widely used measure is $TE_0 = E(\exp(-u)|e)$ built upon prespecified distributions for v and u with $e=v-u$ (Battese and Coelli, 1988; Lee and Tyler, 1978). In our model, TE_0 is not applicable because output is measured in level

while u is an additive unobservable. Furthermore, the distributions of both v and u are eschewed. Since our model accesses the distribution of inefficiency through $E(u|W) = \mu(W;\theta)$, an alternative measure for TE in our model can be constructed as $TE_1 = E(\exp(-u)|W)$. One can employ $\exp(-\mu(W;\theta))$ as a lower bound estimate of TE_1 by Jensen's inequality, or employ $1-\mu(W;\theta)$ as the first order approximation for TE_1. However, TE_1 does not measure TE directly in an usual sense as TE_0 does, a problem similar to recent SF models with distribution-free features (Tran and Tsionas, 2009; Parmeter et al., 2016; Zhou et al., 2020). We refer readers to see detailed discussion about TE measures in Kumbhakar et al. (2020), and left a more direct TE measure in SF models without distribution assumptions for future research.

2.4 Basic assumptions for estimator consistency

In this section, we discuss basic assumptions that are useful to establish the consistency of our proposed estimator for η_{-1} and $A(\cdot)$ in SP-CES-SP. The following notation is used. For any matrix A, define the Frobenius Norm $||A|| = [trace(A^\top A)]^{1/2}$. Let C be a a generic constant with inconsequential magnitude that varies from one place to another. We denote a generic function $\zeta(w) \in C^p$ if $\zeta(w)$ and all of its partial derivatives of order less than or equal to p are continuous and uniformly bounded on its support.

Assumptions:

A1 (1) $\{(Y_{it}, X_{it}, Z_{it}, Q_{it}) : i = 1,\dots,n,\ t = 1,\dots,T\}$ is i.i.d. (identically and independently distributed) across i. Each $X_{jit} > 0$ for $j = 1, ..., d_x$ and the support of X_{jit} is a compact subset of $\Re_+$. (2) The fixed effect is α_i and $E(\alpha_i|X_{it}, Z_{it}, Q_{it}) \neq 0$. (3) v_{it} is $i.i.d.(0, \sigma_v^2)$ across both i and t, independent of other variables. (4) u_{it} is i.i.d. across $i = 1,\dots,n$, $E(u_{it}|\alpha_i, X_{it}, Z_{it}, Q_{it}) = \mu(W_{it};\theta)$, $E(u_{it}^2|\alpha_i, X_i, Z_i, Q_i) < C$ for all i, t with $X_i = \{X_{it}\}_{t=1}^T$. (5) $E(\epsilon_{it}|\alpha_i, X_{it}, Z_{it}, Q_{it}) = 0$ for all i and t with $\epsilon_{it} = v_{it} - (u_{it} - \mu(W_{it};\theta))$.

A2 (1) Denote the true value of $\eta_{-1} = (\gamma_{-1}^\top, \omega_{-1}^\top)^\top$ as η_{-1}^0, which uniquely minimizes the objective function $\mathcal{L}_N(\eta_{-1})$ in (4). For $\Theta_{-1} = \{\Gamma_{-1}, \Omega_{-1}\}$, a compact subset of $\Re^q$ with $q = d_z + d_x + d_w + n - 1$, η_{-1}^0 is contained in the interior of Θ_{-1}. The objective function $\mathcal{L}_N(\eta_{-1})$ is uniquely minimized at η_{-1}^0. (2) $v(\gamma) = \mathbf{Z}^\top\gamma$ is defined on its compact support $\mathcal{V}_\gamma = \{\mathbf{Z}^\top\gamma : \mathbf{Z} \in \mathcal{Z},\ \gamma_{-1} \in \Gamma_{-1}\}$, where $\mathcal{Z}$ is the compact support of $\mathbf{Z}$.

A3 (1) For any $v(\gamma) \in \mathcal{V}_\gamma$ with γ_{-1} in the neighborhood of $\gamma_{-1}^0 \in \Theta_{-1}$, $\exists$ some constant $\delta > 0$ and a $L \times 1$ real vector $\lambda_A \in \Re^L$ such that $\sup_{v(\gamma)\in\mathcal{V}_\gamma} |A(v(\gamma)) - \phi(v(\gamma))^\top\lambda_A| = O(L^{-\delta})$. (2) $L \to \infty$ as $nT \to \infty$. (3) For some nonstochastic sequences ξ_L, $\sup_{v(\gamma)\in\mathcal{V}_\gamma} ||\phi(v(\gamma)|| \leq \xi_L$. (4) as $nT \to \infty$, $\xi_L^2 L/nT \to 0$.

B1 (1) For $v \equiv v(\gamma) \in \mathcal{V}_\gamma$, the kernel function $K(v) : \Re \to \Re$ is symmetric and satisfies (1) $\left|K(v)v^j\right| \leq C$ for all $v \in \Re$ for $j = 0, 1, 2, 3$. (2) $\int |v^j K(v)|dv \leq C$ for $j = 0, 1, 2, 3$. (3)

$\int K(v)dv = 1$, $\int vK(v)dv = 0$, and $\mu_{k,2} = \int v^2 K(v)dv$. (4) $K(v)$ is continuously differentiable on $\Re$ with $|v^j \frac{d}{dv}K(v)| \leq C$ for all $v \in \Re$ and $j = 0, 1, 2, 3$.

B2 (1) $A(\cdot) \in C^2$. (2) For some $\rho > 0$, $E|v_{it}|^{2+\rho} < C$, $E|u_{it}|^{2+\rho} < C$, and $E(|\epsilon_{it}|^{2+\rho}|\alpha_i, X_{it}, Z_{it}, Q_{it}) < C$. (3) $E(\epsilon_{it}^2|\mathbf{Z}_{it}^\top\gamma = \mathbf{z}^\top\gamma) \equiv \sigma_\epsilon^2(\mathbf{z}^\top\gamma) \in C^2$. The density function $0 < f_{v(\gamma)}(\cdot) < C$ of random variable $v(\gamma) = \mathbf{Z}^\top\gamma$ is twice differentiable defined on $\mathcal{V}_\gamma$.

B3 (1) As $n, T \to \infty$, $n/T = \kappa$ for some $0 < \kappa < C$. (2) $\xi_L^2 = O(L)$. (3) $h = O((nT)^{-1/5})$.

Assumptions A1-A3 facilitate our argument on the consistency of PNLS estimator $\widehat{\eta}_{-1}$ in (4). A1(1) focuses on the panel data with large n with i.i.d. assumption. We require large T to avoid incidental parameters problem when handling the fixed effects α_i. A1(2) allows α_i to be arbitrary correlated with (X_{it}, Z_{it}, Q_{it}) for all t. A1(3) imposes a mild condition on the independence of v_{it}. A1(4) assumes that the conditional mean of inefficiency u_{it} is known up to a vector of parameters θ with finite conditional variance. A1(5) specifies zero conditional mean of demeaned composite error for model identification. Assumptions A2(1) ensures the identification on parameters η_{-1}^0, which is necessary to establish the consistency of $\widehat{\eta}_{-1}$ in a similar spirit to the classical assumption of D of Theorem 4.3.1 for the consistency of nonlinear least squares in Amemiya (1985). A2(2) is a common condition for the support of index value as discussed in Ma and Song (2015). A3(1) ensures that, in the first step where parameters are estimated, the series approximation errors on technology function converge to zero as $L \to \infty$, as required in A3(2). A3(3) bounds the magnitudes of the basis functions, and A3(4) gives the rates at which $L \to \infty$ as $nT \to \infty$ such that the asymptotic bias and variance generated in estimating technology are sufficiently small to achieve parametric convergence rate of $\widehat{\eta}_{-1}$.

Assumptions B1-B3 are basic conditions for the consistency of LNK estimator $\widehat{\alpha}(\widehat{\eta})$ in (5). B1 gives conditions on kernel functions, which are frequently used in kernel-based estimation (see, for example, Yao et al. (2018a)). We note that B2 allows a Gaussian kernel to be used. B2(1) assumes $A(\cdot)$ to be sufficiently smooth in order for Taylor expansions to be performed. B2(2) places boundness conditions on the conditional moment of v_{it}, u_{it}, and ϵ_{it} (of order slightly larger than two) to derive asymptotic distribution of LNK under Liapunov's central limit theorem. B2(3) specifically allows conditional heteroscedasticity on ϵ_{it}, and B3(4) assumes that the density of index value is well defined and smooth. B3(1) requires a bounded ratio of n/T when both n and T diverge toward infinity, a condition similar to recent parametric panel model with fixed effects (Fernández-Val and Vidoli, 2016). Specifically, we require a large T relative to n in order to achieve accurate estimation due to the fixed effects estimation. Finally, B3(2) and B3(3) provide expected rates that the number of series basis functions (in the first step) and bandwidths (in the second step) need to be satisfied. See **Remark 4** below for detailed discussion.

With assumptions A-B, our proposed estimator $\widehat{\eta}_{-1}$ and $\widehat{\alpha}(\widehat{\eta})$ are expected to be consistent. This conjecture is made based on the theoretical arguments in Ma and Song (2015), whose regression

model shares some similar features to our SF-CES-SP. Namely, Ma and Song (2015) consider a cross sectional (or pool) varying index coefficient model as $Y_{it} = \sum_{l=1}^{d} m_l(Z_{it}^{\top}\beta_l)X_{lit}^0 + \epsilon_{it}$, where the coefficient of X_{lit}^0 is an unknown function $m_l(\cdot)$, varying with an index $Z_{it}^{\top}\beta_l$ for some coefficient vector β_l. In a particular case, their model is similar to the regression type of SF-CES-SP in (2) when $d = 2$, $m_1(Z_{it}^{\top}\beta_l) \equiv -\mu(W_{it};\theta)$ with $X_{1it}^0 = 1$, and $m_2(Z_{it}^{\top}\gamma) \equiv A(\mathbf{Z}_{it}^{\top}\gamma)$ with $X_{2it}^0 \equiv (\pi_1 X_{1it}^{-\rho} + \cdots + \pi_{d_x} X_{d_x it}^{-\rho})^{-\frac{\nu}{\rho}}$. In other words, our proposed model can be seen as a revised version of Ma and Song (2015)'s model such that $m_1(\cdot)$ is an known function and X_{2it}^0 (with coefficient $m_2(\cdot)$) is not directly observable because it is a scaled index of $X_{1it}, ..., X_{d_x it}$ that depends on $(\pi_1, .., \pi_{d_x}, \rho, \nu)$. Owing to this similarity, Ma and Song (2015)'s model and ours share a common ground where unknown parameters are estimated by approximating unknown functions through series estimation. Therefore, following similar arguments in Ma and Song (2015), together with assumptions A-B above, we expect that our SF-CES-SP can be consistently estimated. Following the insights from both Ma and Song (2015) and Fernández-Val and Vidoli (2016), we further expect that $\widehat{\eta}_{-1}$ converges at a parametric rate of $1/\sqrt{nT}$, and $\widehat{\alpha}(\widehat{\eta})$ at a (univariate) nonparametric rate of $1/\sqrt{nTh}$. We demonstrate below that our conjecture on the consistency and convergence rates are supported by numerical evidence through simulation study.

Remark 4. Our current assumptions do not provide specific guidance on the choice of number of basis function L and bandwidth h. Recall that $N = nT$, the regression optimal order of L satisfies $L = O(N^{\frac{1}{2p+1}})$, obtained by minimizing the integrated squared error of the technology function with p being the degree of smoothness in $A(\cdot)$ (Newey, 1997). When p is low (e.g., $p < 3$), the consistency of $\widehat{\eta}_{-1}$ typically calls for an undersmoothed series estimation on $A(\cdot)$ in order for the convergence rate of $\widehat{\eta}_{-1}$ to be asymptotically unaffected by the series approximation error (Ma and Song, 2015). In such a case, a larger order of L should be considered in estimating $\widehat{\eta}_{-1}$, say $L = O(N^{\frac{1}{2p+1}+\varrho})$ for some $\varrho > 1/(2p(2p+1))$. When p is large (e.g., $p \geq 3$) so that $A(\cdot)$ is sufficiently smooth, we expect a similar result in Ma and Song (2015) that the regression optimal order of L can be implemented in the first step. Given that $\widehat{\eta}_{-1}$ are expected to be $\sqrt{nT}$-consistent, the bandwidth h should satisfy regression optimal order $h = O(N^{-\frac{1}{5}})$ (as given in B3(3)), implying that the data-driven cross-validation can be applied in practice for optimal h.

3 Finite Sample Performance

We provide simulation studies to evaluate the finite sample performance of the proposed estimator. We first illustrate the numerical property of the estimators in (4) for parameters and (5) for the technology function across different choice of T with a fixed n. Second, we investigate whether the performance of the estimators are influenced by different specifications of the inefficiency function $\mu(W_{it};\theta)$. We consider a bivariate production model with two production inputs ($d_x = 2$), and

generate data via

$$Y_{it} = A(Z_{it}^{\top}\gamma_z + \alpha_i)\left(\pi X_{1it}^{-\rho} + (1-\pi)X_{2it}^{-\rho}\right)^{-\frac{\nu}{\rho}} + v_{it} - u_{it}. \tag{6}$$

We specify $d_z = 2$, so $Z_{it} = (Z_{1it}, Z_{2it})$. For $j = 1, 2$, we introduce time dependence by generating $Z_{jit} = Z_{jit}^0 + \zeta_{jit}^z$, where $Z_{jit}^0 \sim U(1,4)$ and $\zeta_{jit}^z = 0.25\zeta_{jit-1}^z + \xi_{it}$ follows an AR(1) process with $\xi_{it} \sim \mathcal{N}(0, 0.25^2)$. We generate $X_{jit} = 0.5(Z_{1it} + Z_{2it}) + \zeta_{jit}^x$ to mimic possible correlations between inputs and environment variables, where $\zeta_{jit}^x = 0.5\zeta_{jit-1}^x + \xi_{it}$. Here, we rescale Z_j and X_j into the range of $[1,4]$ to satisfy assumption A1(1) and A2(2) on the boundness of variables' support. Three specifications of the inefficiency considered. 1) (Scaling) we generate u_{it} with a scaling property $u_{it} = (\sqrt{2/\pi})^{-1}u_i^* \exp(W_{it}^{\top}\theta)$, where $u_i^* \sim |\mathcal{N}(0,1)|$ follows a half-normal distribution with $E(u_i^*|W_{it}) = \sqrt{2/\pi}$, so $\mu(W_{it};\theta) = \exp(W_{it}^{\top}\theta)$; 2) (Log-Probit) we consider a log of probit function in $u_{it} = -\ln(\Phi(W_{it}^{\top}\theta))$ ($\Phi(\cdot)$ being the standard CDF), so $\mu(W_{it};\theta) = -\ln(\Phi(W_{it}^{\top}\theta))$; and 3) (Log-Normal) we specify u_{it} as a log-normal random variable $u_{it} \sim Log\mathcal{N}(0, \sigma^2(W_{it}^{\top}\theta))$, where $\sigma^2(\cdot)$ is the conditional variance function and $\mu(W_{it};\theta) = \exp(0.5W_{it}^{\top}\theta)$. In each case, we draw $v_{it} \sim \mathcal{N}(0, 0.25^2)$, and specify $W_{it}^{\top}\theta = Q_{it}\theta_1 + X_{1it}\theta_2$, where $Q_{it} \sim \mathcal{U}(0,1)$. We construct the fixed effect $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} c_0(\sum_{j=1}^{2}(X_{jit} + Z_{jit}) + Q_{it}) + \xi_i$, where $\xi_i \sim \mathcal{N}(0, 1.25^2)$ and $c_0 = 1$ (i.e., the fixed effects are correlated with the regressors).

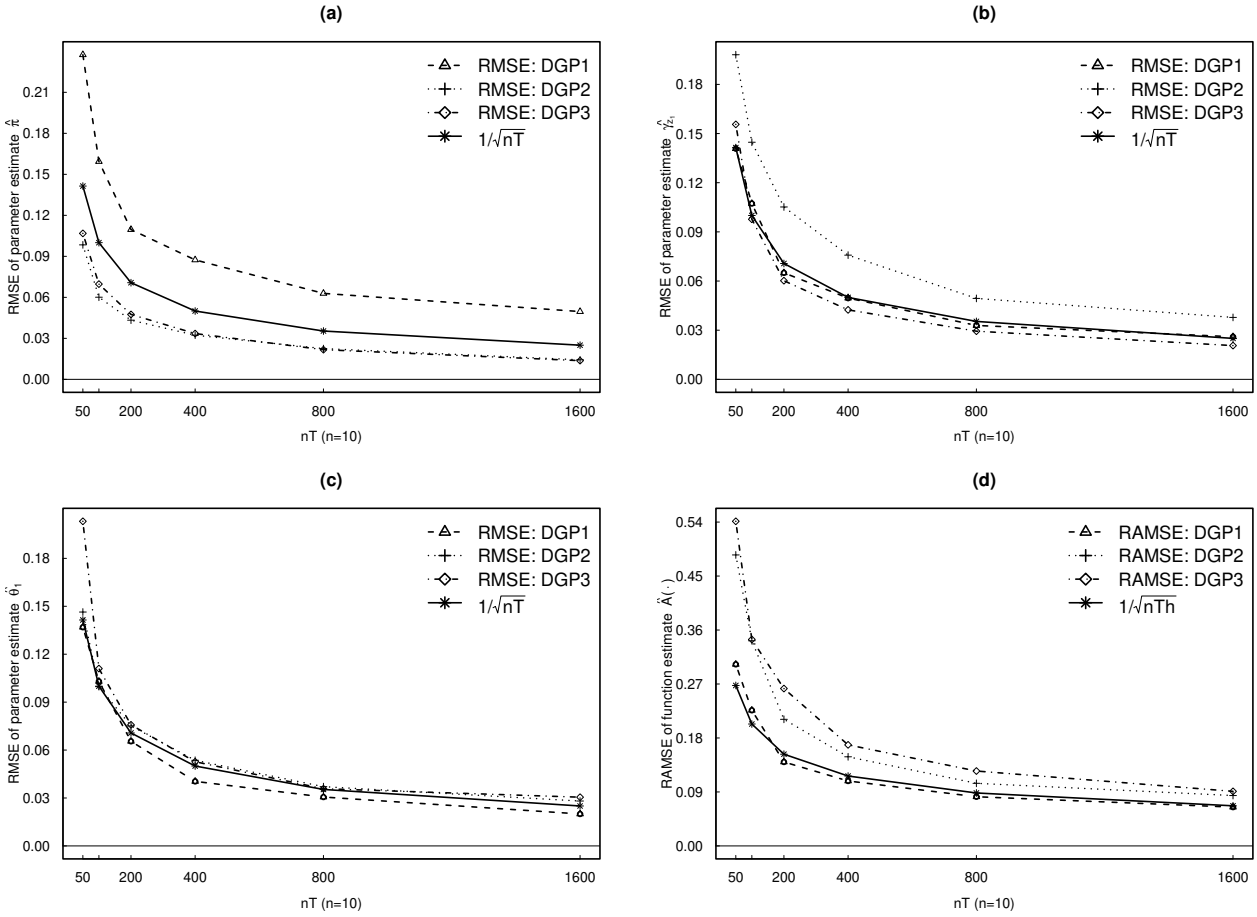
We consider three DGPs with different specifications for $A(\cdot)$, $\gamma = (\gamma_1, \gamma_2, \alpha_{-1})$, and $\omega = (\pi, \rho, \nu, \theta^{\top})$, with $\alpha_{-1} = [\alpha_2, ..., \alpha_n]$ and $\theta = (\theta_1, \theta_2)$. In DGP_1, $A(v) = sin(v\pi/2)$, $\gamma = (2, 4, \alpha_{-1}^{\top})^{\top}$, $\omega = (0.7, 0.5, 0.8, 1.5, -0.5)$; in DGP_2, $A(v) = \sqrt{v\pi}$, $\gamma = (3, 5, \alpha_{-1}^{\top})^{\top}$, $\omega = (0.5, 0, 1, 2, 1)$; and in DGP_3, $A(v) = \exp(v/2)$, $\gamma = (4, 6, \alpha_{-1}^{\top})^{\top}$, and $\omega = (0.3, -0.5, 2, -1, 0.5)$. We rescale γ empirically to satisfy $||\gamma|| = 1$. The values of ρ and ν are chosen to reflect a production function with complementary elasticity of substitution ($\rho > 0$) and decreasing return to scale ($\nu < 1$) in DGP_1; unit elasticity of substitution ($\rho = 0$) and constant return to scale ($\nu = 1$) in DGP_2; and substitutable elasticity of substitution ($\rho < 0$) and increasing return to scale ($\nu > 1$) in DGP_3.

In our first step estimation, we implement cubic B-spline basis functions (of order $m = 3$) in (4) and place a total of $L = J + m + 1$ knots on the support of $v_{it}(\gamma)$, where J refers to the number of (strictly) interior knots of $v_{it}(\gamma)$. Since the function $A(\cdot)$ under each DGP is sufficiently smooth, we follow the arguments in **Remark 3** to select $J = [(nT)^{1/5}]$ as a *rule-of-thumb* (ROT) approach, where $[v]$ is the integer part of a real number v. We place the l^{th} interior knot on the $(l/J+1)^{th}$ percentile of $v_{it}(\gamma)$ for $l = 1, ..., J$. In the second step, we use a second-order Gaussian kernel function in (5) with a ROT bandwidth $h_j = \widehat{\sigma}_j(nT)^{-1/5}$ satisfying B3(3), where $\widehat{\sigma}_j$ is the standard deviation of $\{v_{it}(\widehat{\gamma})\}_{i=1,t=1}^{n,T}$. Throughout the experiment, we fix $n = 10$, set $T = (10, 20, 40)$, and conduct 500 repetitions. We evaluate the performance of the PNLS estimator $\widehat{\gamma}$ by root mean squared error (RMSE), (absolute) bias (BIAS), and standard deviation (STD). We evaluate the performance of LNK estimator $\widehat{A}(\cdot)$ via root average MSE (RAMSE), average BIAS (ABIAS), and

Table 1: Simulation Results for SF-CES-SP with Scaling Inefficiency in DGP_{1-3}

	(n,T)		RMSE	BIAS	STD		RAMSE	ABIAS	ASTD
DGP_1	(10,10)	$\widehat{\gamma}_{z_1}$	0.1032	0.0820	0.0630	$\widehat{A}$	0.2056	0.1568	0.1931
		$\widehat{\pi}$	0.1393	0.1079	0.0886				
		$\widehat{\rho}$	0.0687	0.0550	0.0413				
		$\widehat{\nu}$	0.1426	0.1031	0.0989				
		$\widehat{\theta}_1$	0.0883	0.0701	0.0539				
	(10,20)	$\widehat{\gamma}_{z_1}$	0.0616	0.0474	0.0396	$\widehat{A}$	0.1505	0.1173	0.1376
		$\widehat{\pi}$	0.1127	0.0903	0.0677				
		$\widehat{\rho}$	0.0423	0.0330	0.0266				
		$\widehat{\nu}$	0.0995	0.0795	0.0602				
		$\widehat{\theta}_1$	0.0583	0.0448	0.0375				
	(10,40)	$\widehat{\gamma}_{z_1}$	0.0432	0.0350	0.0254	$\widehat{A}$	0.1044	0.0868	0.0906
		$\widehat{\pi}$	0.0823	0.0638	0.0524				
		$\widehat{\rho}$	0.0300	0.0244	0.0176				
		$\widehat{\nu}$	0.0685	0.0562	0.0395				
		$\widehat{\theta}_1$	0.0449	0.0342	0.0292				
DGP_2	(10,10)	$\widehat{\gamma}_{z_1}$	0.1463	0.1105	0.0964	$\widehat{A}$	0.2948	0.2092	0.2774
		$\widehat{\pi}$	0.0653	0.0506	0.0415				
		$\widehat{\rho}$	0.1043	0.0781	0.0695				
		$\widehat{\nu}$	0.1192	0.0987	0.0673				
		$\widehat{\theta}_1$	0.0996	0.0749	0.0659				
	(10,20)	$\widehat{\gamma}_{z_1}$	0.1091	0.0871	0.0660	$\widehat{A}$	0.1992	0.1550	0.1812
		$\widehat{\pi}$	0.0465	0.0354	0.0303				
		$\widehat{\rho}$	0.0848	0.0658	0.0537				
		$\widehat{\nu}$	0.0797	0.0666	0.0439				
		$\widehat{\theta}_1$	0.0704	0.0569	0.0417				
	(10,40)	$\widehat{\gamma}_{z_1}$	0.0881	0.0726	0.0502	$\widehat{A}$	0.1381	0.1075	0.1326
		$\widehat{\pi}$	0.0333	0.0269	0.0198				
		$\widehat{\rho}$	0.0679	0.0542	0.0411				
		$\widehat{\nu}$	0.0596	0.0479	0.0357				
		$\widehat{\theta}_1$	0.0523	0.0420	0.0313				
DGP_3	(10,10)	$\widehat{\gamma}_{z_1}$	0.0781	0.0653	0.0432	$\widehat{A}$	0.3466	0.2625	0.3042
		$\widehat{\pi}$	0.0631	0.0505	0.0380				
		$\widehat{\rho}$	0.0964	0.0798	0.0544				
		$\widehat{\nu}$	0.1258	0.1026	0.0731				
		$\widehat{\theta}_1$	0.0951	0.0731	0.0611				
	(10,20)	$\widehat{\gamma}_{z_1}$	0.0580	0.0463	0.0350	$\widehat{A}$	0.2688	0.2040	0.2381
		$\widehat{\pi}$	0.0417	0.0337	0.0246				
		$\widehat{\rho}$	0.0718	0.0563	0.0449				
		$\widehat{\nu}$	0.0943	0.0796	0.0510				
		$\widehat{\theta}_1$	0.0804	0.0653	0.0471				
	(10,40)	$\widehat{\gamma}_{z_1}$	0.0415	0.0327	0.0258	$\widehat{A}$	0.1984	0.1435	0.1746
		$\widehat{\pi}$	0.0323	0.0265	0.0187				
		$\widehat{\rho}$	0.0517	0.0395	0.0335				
		$\widehat{\nu}$	0.0721	0.0594	0.0411				
		$\widehat{\theta}_1$	0.0566	0.0433	0.0366				

Figure 1: Convergence Rates of Parameter (a)-(c) and Function (d) Estimators in DGP_{1-3}



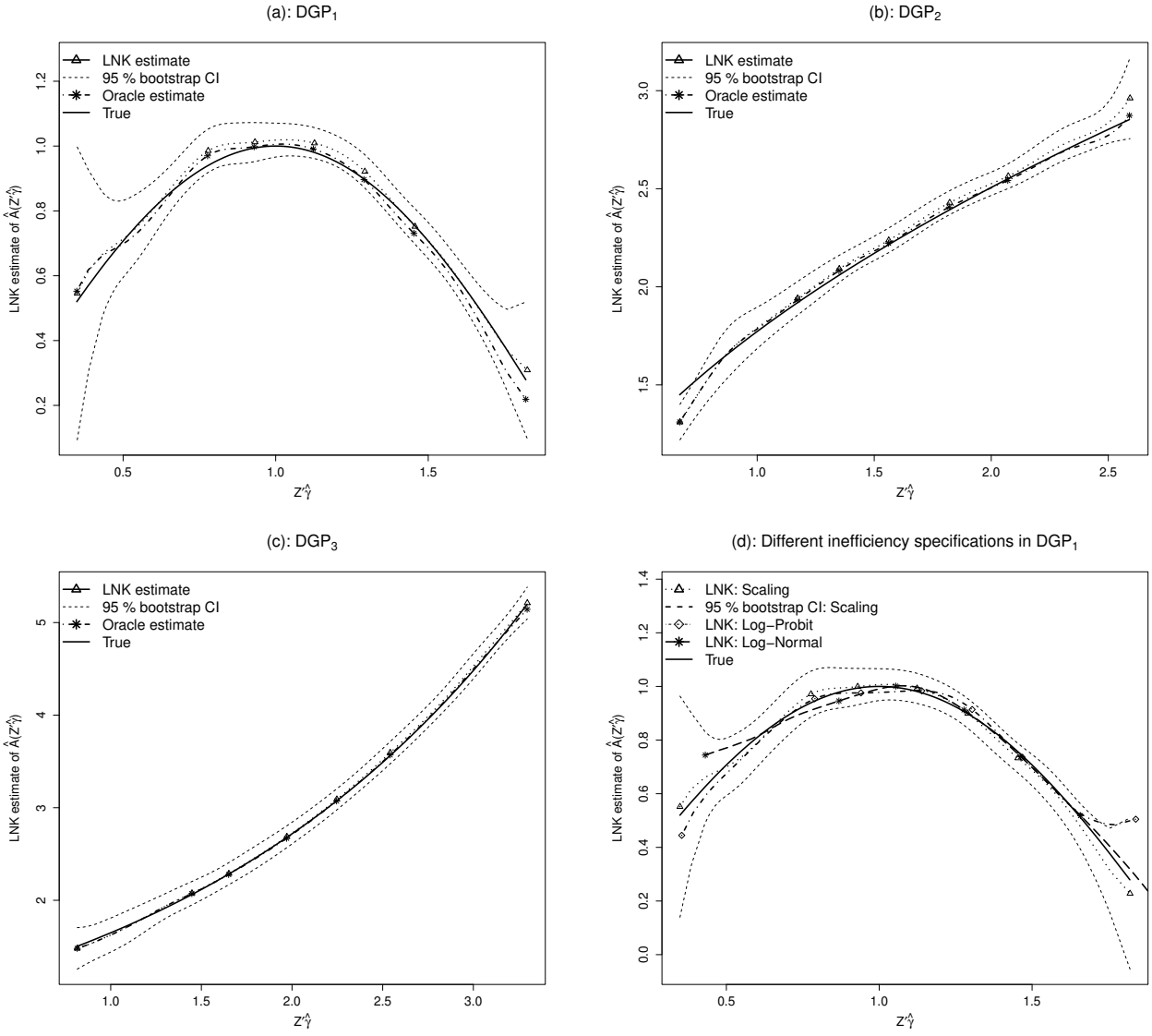
average STD (ASTD).

Table 1 summarizes simulation results for our SF-CES-SP estimator by considering a scaling distribution of inefficiency in DGP_{1-3}. Given that the performance of each coefficient estimate in $\widehat{\gamma}_z$ and $\widehat{\theta}$ is similar, we only report the statistic measures for $\widehat{\gamma}_{z_1}$ and $\widehat{\theta}_1$ to save space. Conditioning on each sample, the parametric estimator uniformly outperforms the nonparametric estimator (as expected) with different DGPs. Both parameter and function estimator are consistent in that all measures decay toward zero as T doubles regardless of the DGP. More specifically, the estimators converge at their expected rates as our conjecture in Section 2.4. To clearly demonstrate the convergence rates, we re-estimate our model in (6) by fixing $n = 10$, raising $T = (5, 10, 20, 40, 80, 160)$, and compute RMSE and RAMSE given each sample size. Figure 1 plots the RMSE of one frontier parameter $\widehat{\pi}$ in panel (a), one technology parameter $\widehat{\gamma}_{z_1}$ in panel (b), and one inefficiency parameter $\widehat{\theta}_1$ in panel (c). In each panel, the RMSE under DGP_1 (dashed line with $\triangle$), DGP_2 (dot line with $+$), and DGP_3 (dot-dashed line with $\diamond$) is plotted against the parametric rate of $1/\sqrt{nT}$ (solid line

Table 2: Simulation Results for SF-CES-SP with Different Distribution of Inefficiency in DGP_1

	(n,T)		RMSE	BIAS	STD		RAMSE	ABIAS	ASTD
Scaling	(10,10)	$\widehat{\gamma}_1$	0.1071	0.0822	0.0629	$\widehat{A}$	0.2086	0.1565	0.1915
		$\widehat{\pi}$	0.1376	0.1085	0.0887				
		$\widehat{\rho}$	0.0679	0.0561	0.0408				
		$\widehat{\nu}$	0.1431	0.1010	0.1006				
		$\widehat{\theta}_1$	0.0919	0.0710	0.0536				
	(10,20)	$\widehat{\gamma}_1$	0.0601	0.0482	0.0394	$\widehat{A}$	0.1511	0.1207	0.1407
		$\widehat{\pi}$	0.1152	0.0884	0.0664				
		$\widehat{\rho}$	0.0428	0.0339	0.0275				
		$\widehat{\nu}$	0.0979	0.0778	0.0619				
		$\widehat{\theta}_1$	0.0586	0.0442	0.0368				
	(10,40)	$\widehat{\gamma}_1$	0.0432	0.0332	0.0252	$\widehat{A}$	0.1051	0.0854	0.0904
		$\widehat{\pi}$	0.0830	0.0643	0.0532				
		$\widehat{\rho}$	0.0306	0.0254	0.0180				
		$\widehat{\nu}$	0.0684	0.0561	0.0393				
		$\widehat{\theta}_1$	0.0456	0.0336	0.0290				
Log-Probit	(10,10)	$\widehat{\gamma}_1$	0.0951	0.0723	0.0621	$\widehat{A}$	0.2705	0.2147	0.2611
		$\widehat{\pi}$	0.0974	0.0781	0.0585				
		$\widehat{\rho}$	0.1428	0.1072	0.0947				
		$\widehat{\nu}$	0.1526	0.1238	0.0896				
		$\widehat{\theta}_1$	0.1508	0.1193	0.0927				
	(10,20)	$\widehat{\gamma}_1$	0.0589	0.0474	0.0351	$\widehat{A}$	0.2113	0.1598	0.1819
		$\widehat{\pi}$	0.0688	0.0575	0.0380				
		$\widehat{\rho}$	0.1068	0.0838	0.0688				
		$\widehat{\nu}$	0.0995	0.0766	0.0638				
		$\widehat{\theta}_1$	0.0911	0.0684	0.0606				
	(10,40)	$\widehat{\gamma}_1$	0.0425	0.0339	0.0257	$\widehat{A}$	0.1812	0.1407	0.1330
		$\widehat{\pi}$	0.0577	0.0476	0.0328				
		$\widehat{\rho}$	0.0688	0.0458	0.0381				
		$\widehat{\nu}$	0.0699	0.0558	0.0424				
		$\widehat{\theta}_1$	0.0656	0.0523	0.0399				
Log-Normal	(10,10)	$\widehat{\gamma}_1$	0.1423	0.1107	0.0899	$\widehat{A}$	0.3228	0.2499	0.3077
		$\widehat{\pi}$	0.1267	0.1039	0.0729				
		$\widehat{\rho}$	0.1000	0.0781	0.0628				
		$\widehat{\nu}$	0.1722	0.1305	0.1129				
		$\widehat{\theta}_1$	0.0322	0.0251	0.0203				
	(10,20)	$\widehat{\gamma}_1$	0.1083	0.0886	0.0627	$\widehat{A}$	0.2704	0.1997	0.2346
		$\widehat{\pi}$	0.0721	0.0579	0.0432				
		$\widehat{\rho}$	0.0745	0.0618	0.0420				
		$\widehat{\nu}$	0.1116	0.0937	0.0609				
		$\widehat{\theta}_1$	0.0269	0.0209	0.0170				
	(10,40)	$\widehat{\gamma}_1$	0.0702	0.0595	0.0374	$\widehat{A}$	0.2216	0.1643	0.1664
		$\widehat{\pi}$	0.0639	0.0470	0.0334				
		$\widehat{\rho}$	0.0513	0.0486	0.0310				
		$\widehat{\nu}$	0.0791	0.0587	0.0533				
		$\widehat{\theta}_1$	0.0205	0.0161	0.0127				

Figure 2: Technology Function Estimation (LNK) with $(n, T) = (20, 50)$



with $*$). Clearly, each parameter estimate decays at a rate quite close to the theoretical rate. Similar observation is made on RAMSE of $\widehat{A}(\cdot)$, which closely resembles the univariate nonparametric rate of $1/\sqrt{nTh}$ in panel (d) of Figure 1. The numerical properties of our estimator support our conjecture that the parameter (function) estimator is $\sqrt{nT}$ ($\sqrt{nTh}$)-consistent.

To clearly demonstrate the performance of function estimator, Figure 2 plots LNK estimates $\widehat{A}(\cdot)$ with a large sample $(n, T) = (20, 50)$ in Panels (a)-(c) for DGP_1, DGP_2, and DGP_3, respectively. Each LNK function estimate (dot line with $\triangle$) is plotted against its point-wise 95% wild bootstrap confidence interval (CI) with 399 repetitions (dashed line), true function (solid line), and oracle

estimates in which η is assumed to be known (dot-dashed line with $*$). Consistent with Table 1, the estimates sufficiently reveal the shape of each function, and closely resemble the corresponding oracle estimates and true functions across all DGPs. We note that the bias is smaller for functions with a lower degree of curvature (i.e., DGP_1).

Table 2 investigates the performance of our proposed estimator under three different specifications for $\mu(W_{it};\theta)$. We report the estimation results with different sample sizes in DGP_1 under Scaling (upper panel), Log-Probit (middle panel), and Log-Normal (lower panel). The performance of both parameters and function estimator are relatively similar under the cases of Scaling and Log-Probit, but slightly worsen in the case of Log-Normal when the sample size is moderate. The estimator is qualitatively similar across all three specifications in $\mu(W_{it};\theta)$, again illustrating its consistency property. Using the sample size $(n,T) = (20,50)$, Figure 2 plots in panel (d) the function estimates under Scaling (dot line with $\triangle$) against these under Log-Probit (dot short-dashed line with $\diamond$) and Log-Normal (dot long-dashed line with $*$), together with 95% bootstrap CI under the Scaling case and true function. It is clear that the estimation difference under different specifications of inefficiency function becomes negligible with a larger sample size. Overall, our proposed estimator demonstrates appealing numerical performance in estimating unknown parameters and functions in SF-CES-SP, which is fairly robust to different specifications of the inefficiency.

4 Empirical Application

A number of studies have investigated country productive efficiency through SF models. For instance, Iyer et al. (2008) evaluate the roles of trade variables among developed (OECD) countries. Yao et al. (2018a) and Yao et al. (2018b) focus specifically on the effect of human capital among both developed and developing countries. The SF models adopted by the aforementioned studies are similar in three aspects that could be potentially restrictive for the efficiency analysis. First, the frontier is specified as logged CD with either parametric or semiparametric structure. The implied unit elasticity of inputs substitution may not necessarily hold in countries with different technology bias and, as shown below, applying log transformation on a SF model would bias estimation on both CD frontier and inefficiency. Second, the composite error is assumed to follow specific distributions for the purpose of model identification, but the justification of correct distributional assumptions can be difficult with a real dataset. Finally, the unobserved heterogeneities (or fixed effects) in countries are not controlled, which can lead to inconsistent model estimation due to their potential correlation with covariates appearing in the frontier and/or inefficiency.

We are therefore motivated to illustrate our semiparametric SF model with flexible features on both the frontier and inefficiency while properly handling country fixed effects. Following the insights from our simulation results, we are encouraged to employ a panel dataset with small n and large T. To this end, we narrow our scope to consider OECD countries with membership

Table 3: Data Descriptive Summary Statistics

Variables	Description	Mean	SD	Min	Max
Y	Output-side real GDP	1285441	2787333	3647.29	20860506
K	Capital stock	5335775	10301105	18818.94	69059088
HL	Human capital-augmented labor, the product of human capital index (HC) and number of persons engagaed (L)	51.54	97.72	0.20	593.52
EXS	Share of merchandise exports	0.37	0.25	0.02	1.39
$FDI\text{-}INS$	FDI: Inward flows and stocks	0.03	0.09	-0.66	0.83
$FDI\text{-}OUTS$	FDI: Outward flows and stocks	0.05	0.16	-0.38	1.59
(n,T)	(20,50)				

Note: Variables Y, K, HL, and EXS are obtained from the tenth version of PWT, measured in millions at 2017 US dollars. Variables $FDI\text{-}INS$ and $FDI\text{-}OUTS$ are constructed using FDI inward and outward flows (in millions) from UNCTAD, which are adjusted at 2017 US dollar and divided by Y. All data are collected for 20 OECD countries during 1970-2019.

since 1962 in North America and Europe from 1970 to 2019. This gives us 20 countries over five decades, so $n = 20$ and $T = 50$ (i.e., equivalent to the sample size used to generate Figure 2). We source our data from the tenth version of Penn World Table (PWT) and United Nations Conference on Trade and Development (UNCTAD), which are by far the most complete panel data sources with aggregate production and trade variables at the country level. Using PWT, we collect data on output Y as the output side real GDP in millions; two inputs K as real capital stock in millions and HL as the human capital augmented labor;[8] and one environment variable EXS as export share of a country's real GDP, which is likely to generate non-trivial impacts on technology (Malikov et al., 2020). Given that inefficiency plays an important role in influencing productivity change (Kumbhakar and Lovell, 2000), we further include EXS as one inefficiency determinant to capture a possible indirect effect of EXS on technology through affecting productive inefficiency. Using UNCTAD, we follow the arguments by Iyer et al. (2008) to collect the share of foreign direct investment inward flows ($FDI\text{-}INS$) and outward flows ($FDI\text{-}OUTS$) relative to real GDP as two additional inefficiency determinants. All variables (except for HL) are adjusted at 2017 US dollars to facilitate estimation comparison across countries and time.

In sum, we perform production efficiency analysis among OECD countries through our SF-CES-SP (in regression form (2)) as

$$Y_{it} = A\left(EXS_{it}\gamma_{exs} + \alpha_i\right)\left[\pi K_{it}^{-\rho} + (1-\pi)HL_{it}^{-\rho}\right]^{-\frac{\nu}{\rho}} - \mu\left(W_{it};\theta\right) + \epsilon_{it}, \tag{7}$$

where $i = 1, \ldots, 20$ is the OECD country index, $t = 1, \ldots, 50$ represents years from 1970 to 2019, and $\epsilon_{it} = v_{it} - \left(u_{it} - \mu\left(W_{it};\theta\right)\right)$. Here, we have inputs $X = (K, HL)$, environment variables

[8] We follow Henderson and Russell (2005) to construct $HL = HC \times L$, where HC is the human capital index based on years of schooling and returns to education from PWT, and L the millions of persons engaged in employment.

Z with EXS plus $n-1=19$ dummies (controlling for country fixed effects), and inefficiency determinants $W=(Cont, EXS, FDI\text{-}INS, FDI\text{-}OUTS)$, where $Cont$ is a constant term. Table 3 summarizes the distribution of each variable by reporting their corresponding average (Mean), standard deviation (SD), minimum (Min), and maximum (Max). We highlight that the USA exhibits the highest level of Y, K, and HL. Belgium has the largest EXS and Luxembourg tops the list of both FDI variables. In contrast, Iceland and Turkey show the lowest values in production variables (Y, K, HL) and trade variables $(EXS, FDI\text{-}INS, FDI\text{-}OUTS)$, respectively.

To facilitate comparison between our SF-CES-SP and conventional models in the SF literature, we further estimate four benchmark parametric SF panel models with different frontier structures. The first benchmark model considers a parametric CD frontier in level (SF-CD-P-Level) as

$$Y_{it}=A_p\left(EXS_{it}\gamma_{exs}+\alpha_i\right)K_{it}^{\beta_K}HL_{it}^{\beta_{HL}}-\mu\left(W_{it};\theta\right)+\epsilon_{it}, \tag{8}$$

where the technology function $A_p(v)=\exp(v)$ has a known exponential structure, and parameters β_K and β_{HL} are the output elasticity of K and HL, respectively. The second benchmark model considers a log-transformed CD frontier (SF-CD-P-Log), which has been commonly employed in applied works. Compared to SF-CD-P-Level, the structure of SF-CD-P-Log is more delicate due to logging in the presence of productive inefficiency. To see this, we define $F_{it}\equiv A_p\left(EXS_{it}\gamma_{exs}+\alpha_i\right)K_{it}^{\beta_K}HL_{it}^{\beta_{HL}}$ as the frontier and $\zeta_{it}=1+\left(v_{it}-u_{it}\right)/F_{it}$ as an error term, so (8) is equivalently to $Y_{it}=F_{it}\zeta_{it}$. Since $\ln\left(Y_{it}\right)=\ln\left(F_{it}\right)+\ln\left(\zeta_{it}\right)$, a first-order Taylor expansion on $\ln\left(\zeta_{it}\right)$ gives an approximation $\ln\left(1+\left(v_{it}-u_{it}\right)/F_{it}\right)\approx\frac{v_{it}-u_{it}}{F_{it}}-\frac{\left(v_{it}-u_{it}\right)^2}{2F_{it}^2}=-\frac{\mu(W_{it};\theta)}{F_{it}}+\frac{\epsilon_{it}}{F_{it}}-\frac{\left(v_{it}-u_{it}\right)^2}{2F_{it}^2}$. Thus, the explicit form of SF-CD-P-Log is

$$\ln\left(Y_{it}\right)=EXS_{it}\gamma_{exs}+\alpha_i+\ln\left(K_{it}\right)\beta_K+\ln\left(HL_{it}\right)\beta_{HL}-\mu^*\left(W_{it};\theta\right)+\epsilon_{it}^*, \tag{9}$$

where $\mu^*\left(W_{it};\theta\right)=\mu\left(W_{it};\theta\right)/F_{it}$ and $\epsilon_{it}^*=\frac{\epsilon_{it}}{F_{it}}-\frac{\left(v_{it}-u_{it}\right)^2}{2F_{it}^2}$. As a result, the log transformation on (8) introduces non-negligible estimation bias on parameters in (9) because $\mu^*\left(W_{it};\theta\right)\neq\mu\left(W_{it};\theta\right)$ and $E\left(\epsilon_{it}^*|EXS_{it},\alpha_i,K_{it},HL_{it},W_{it}\right)\neq 0$ in general. The third benchmark model specifies a CES frontier in level (SF-CES-P-Level) as

$$Y_{it}=A_p\left(EXS_{it}\gamma_{exs}+\alpha_i\right)\left[\pi K_{it}^{-\rho}+(1-\pi)HL_{it}^{-\rho}\right]^{-\frac{\nu}{\rho}}-\mu\left(W_{it};\theta\right)+\epsilon_{it}, \tag{10}$$

which differs from SF-CES-SP in (7) in terms of the exponential form of technology. The final benchmark model is the log-transformed version of SF-CES-P-Level through Kmenta's approximation (SF-CES-P-Log). Namely, let the frontier be $F_{it}=A_p\left(EXS_{it}\gamma_{exs}+\alpha_i\right)f_{it}(\rho)$, with $f_{it}(\rho)=\left[\pi K_{it}^{-\rho}+(1-\pi)HL_{it}^{-\rho}\right]^{-\frac{\nu}{\rho}}$. Since $\ln\left(F_{it}\right)=EXS_{it}\gamma_{exs}+\alpha_i+\ln\left(f_{it}(\rho)\right)$, a first-order Taylor expansion on $\ln\left(f_{it}(\rho)\right)$ around $\rho=0$ gives an approximation $\ln\left(f_{it}(\rho)\right)\approx\ln\left(K_{it}\right)b_1+\ln\left(HL_{it}\right)b_2+\left[\ln\left(K_{it}\right)-\ln\left(HL_{it}\right)\right]^2b_3$, where $b_1=\nu\pi, b_2=\nu(1-\pi)$, and $b_3=-\nu\rho(1-\pi)$. Our interested pa-

Table 4: Summary of Estimation Results in Frontier

Panel A	SF-CES-SP		SF-CD-P-Level		SF-CD-P-Log		SF-CES-P-Level		SF-CES-P-Log	
$\widehat{\pi}$	0.6730***	(0.1487)	-	-			0.6889***	(0.1425)	0.2689**	(0.1324)
$\widehat{\rho}$	0.0441***	(0.0103)	-	-			0.0870***	(0.0097)	-0.0967***	(0.0325)
$\widehat{\nu}$	1.1432***	(0.0540)	-	-			1.3587***	(0.0514)	1.0225***	(0.0740)
$\widehat{\beta}_K$	-	-	0.6061***	(0.0118)	0.5003***	(0.0116)	-	-	-	-
$\widehat{\beta}_{HL}$	-	-	0.5143***	(0.0124)	0.5217***	(0.0120)	-	-	-	-
Panel B	Elas(K)	Elas(HL)	Elas(K)	Elas(HL)	Elas(K)	Elas(HL)	Elas(K)	Elas(HL)	Elas(K)	Elas(HL)
0.10	0.5766	0.4675	-	-	-	-	0.5073	0.4517	0.5006	0.3380
0.25	0.5856	0.4707	-	-	-	-	0.5137	0.4578	0.5123	0.3447
0.50	0.5889	0.4755	-	-	-	-	0.5196	0.4638	0.5350	0.3529
0.75	0.5940	0.4806	-	-	-	-	0.5277	0.4696	0.5427	0.3606
0.90	0.5994	0.4848	-	-	-	-	0.5332	0.4768	0.5629	0.3698
Mean	0.5889	0.4757	0.6061	0.5143	0.5003	0.5217	0.5201	0.4640	0.5326	0.3532
Panel C	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
RTS	1.0646	1.0641	1.1204	1.1204	1.0220	1.0220	1.3429	1.3382	0.9841	0.9842

Note: Standard errors are reported in parenthesis beneath the estimated coefficients in Panel A. ***,**,* stands for the statistically significant level at 1%, 5% and 10%, respectively.

rameters in the frontier can be readily recovered as $\nu = b_1 + b_2, \pi = b_1/\nu$, and $\rho = b_3\nu/b_1b_2$. Thus, taking log on both sides gives an explicit form of SF-CES-P-Log as

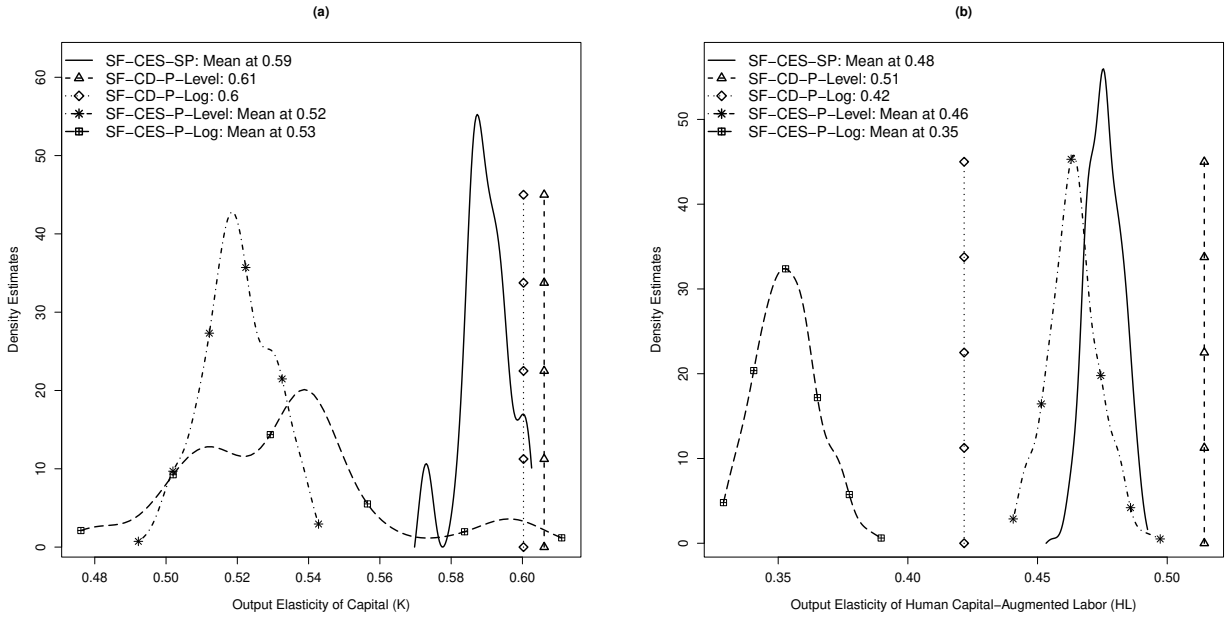
$$\ln(Y_{it}) = EXS_{it}\gamma_{exs} + \alpha_i + \ln(K_{it})\, b_1 + \ln(HL_{it})\, b_2 + [\ln(K_{it}) - \ln(HL_{it})]^2\, b_3 - \mu^*(W_{it};\theta) + \epsilon^*_{it}, \quad (11)$$

where $\mu^*(W_{it};\theta)$ and ϵ^*_{it} are defined similarly as in (9), except that F_{it} now takes CES form and, more importantly, ϵ^*_{it} now contains Taylor's expansion error not only from $\ln(\zeta_{it})$ but also $\ln(f_{it}(\rho))$. Due to this reason, we expect to observe severer estimation bias from SF-CES-P-Log than that from SF-CD-P-Log. As discussed below, our conjecture is indeed supported by our empirical results.

To implement SF-CES-SP, we follow the insights from our simulation study to employ ROT interior knots $J = [(nT)^{1/5}]$ for estimating η_{-1} and cross-validation bandwidth h for estimating $A(\cdot)$. To provide a fair comparison, all four benchmark models (8)-(11) are estimated without imposing distributional assumptions on v_{it} and u_{it}. Given that the estimation results remain qualitatively similar when different functional forms of $\mu(W_{it};\theta)$ are considered, we report the empirical results below for $\mu(W_{it};\theta) = \exp\left(W_{it}^\top\theta\right)$.

We first compare several important metrics regarding the production frontier across different models. *Panel A* of Table 4 reports estimated parameters $(\widehat{\pi}, \widehat{\rho}, \widehat{\nu})$ in models with CES frontiers and parameters $(\widehat{\beta}_K, \widehat{\beta}_{HL})$ in models with CD frontiers. All parameters estimates (except for $\widehat{\pi}$ in SF-CES-P-Log) are significant at the 1% level. Our SF-CES-SP reveals a production function with about 2/3 uses of K as indicated by $\widehat{\pi} = 0.673$, consistent with conventional wisdom that developed countries are commonly capital intensive. The estimate of $\widehat{\rho}$ is 0.044, suggesting that K and HL

Figure 3: Density of Output Elasticity of Capital (a) and Augmented Labor (b)



in OECD countries are complementary given that the implied elasticity of input substitution is $\widehat{\sigma} = 1/(1+\widehat{\rho}) = 0.958$. The underlying production is also suggested to exhibit increasing returns to scale (RTS) as indicated by $\widehat{\nu} = 1.143$, significantly larger than unity at the 1% level. Compared to SF-CES-SP, parameters (π, ρ) are slightly overestimated under SF-CES-P-Level but largely underestimated under SF-CES-P-Log which, in particular, suggests that K accounts for less than 1/3 of total inputs ($\widehat{\pi} = 0.269$). In contrast to SF-CES-SP, both parametric CES models suggest a substitutable relationship between K and HL, with $\widehat{\sigma} = 1.095$ in SF-CES-PLevel and $\widehat{\sigma} = 1.107$ in SF-CES-P-Log. We highlight that the estimate $\widehat{\rho}$ across all three SF models with CES frontier significantly differs from zero at the 1% level, indicating that a CD production frontier is restrictive in our study. Finally, the constant estimate for RTS ($\widehat{\nu}$) is relatively over (under) estimated in SF-CES-P-Level (SF-CES-P-Log).

With the parameter estimates above, we compute the estimated output elasticity (i.e., the share) of K (Elas(K)) and HL (Elas(HL)) in models with CES frontier, where

$$\text{Elas}(K) = \frac{\widehat{\pi} K^{-\widehat{\rho}}}{\left[\widehat{\pi} K^{-\widehat{\rho}} + (1-\widehat{\pi}) HL^{-\widehat{\rho}}\right]^{\widehat{\nu}}}, \quad \text{Elas}(HL) = \frac{(1-\widehat{\pi}) HL^{-\widehat{\rho}}}{\left[\widehat{\pi} K^{-\widehat{\rho}} + (1-\widehat{\pi}) HL^{-\widehat{\rho}}\right]^{\widehat{\nu}}}.$$

Clearly, the output elasticity of each input is country-year specific under CES as opposed to constants $(\widehat{\beta}_K, \widehat{\beta}_{HL})$ under CD. *Panel B* of Table 4 reports Elas(K) and Elas(HL) evaluated at five different percentiles $(0.1, 0.25, 0.5, 0.75, 0.9)$ and the mean. In SF-CES-SP, the mean of output elasticity of K (HL) takes 0.59 (0.48), which is lower than the conventional value of 2/3 (1/3) possibly

due to the effect of human capital through HL. Compared to SF-CES-SP, both SF-CES-P-Level and SF-CES-P-Log yield a lower Elas(K) with means of 0.52 and 0.53, respectively. Although Elas(HL) remains quantitatively similar under SF-CES-P-Level with a mean of 0.46, the log approximation on CES leads to a significant reduction on Elas(HL) in SF-CES-P-Log, with a mean of 0.35. Across all models, the largest Elas(K) and Elas(HL) stem from SF-CD-P-Level with 0.61 and 0.51, respectively. When the log-transformation is applied, we note that SF-CD-P-Log discloses a more even share of two inputs.

Figure 3 provides a more vivid picture of the dispersion of Elas(K) and Elas(HL) across different models in panels (a) and (b), respectively. In each panel, we plot the kernel density of elasticity of K or HL by SF-CES-SP (solid curve) against those by SF-CD-P-Level (vertical dashed line with $\triangle$), SF-CD-P-Log (vertical dot line with $\diamond$), SF-CES-P-Level (dot-dashed curve with $*$), and SF-CES-P-Log (long-dashed curve with $\boxplus$). We find that the nonparametric technology under SF-CES-SP makes the magnitude of Elas(HL) roughly similar to, yet Elas(K) quite distinguishable from, its parametric counterparts by SF-CES-P-Level. Furthermore, the log-approximation on CES results in a notably larger variation in Elas(K) and smaller magnitude in Elas(HL) under SF-CES-P-Log. The impact of log-transformation can also be observed from SF models with CD, from which the Elas (HL) differs significantly across SF-CD-P-Level and SF-CD-P-Log. As a result, the large difference in estimated elasticity of inputs across different models constitutes different estimates of (observation specific) return to scale as RTS=Elas(K)+Elas(HL). *Panel C* of Table 4 reports the mean and median of the RTS, which imply a production function with increasing RTS under SF-CES-SP, SF-CES-P-Level, and SF-CD-P-Level; constant RTS under SF-CD-Log; and decreasing RTS under SF-CES-P-Log.[9]

We next compare several results regarding the technology function. *Panel A* of Table 5 reports estimated coefficients of EXS (i.e., $\widehat{\gamma}_{exs}$) in the technology function across different SF models. By controlling for country fixed effects, we observe a significant positive impact of export share on technology at the 1% level across all models, albeit in different magnitudes. The estimate is given by 0.671 under SF-CES-SP (assuming unknown structure of $A(\cdot)$), about 12% lower than the estimate of 0.751 under SF-CES-P-Level (assuming known structure of $A_p(\cdot)$ being exponential). However, applying log-transformation on CES significantly reduces the magnitude of estimate toward 0.058 under SF-CES-P-Log. A similar observation is made on SF models with CD, where the estimate 0.367 in SF-CD-P-Level drops by almost three-fold to 0.092 in SF-CD-P-Log.

To evaluate how countries' export share evolve technology in the presence of latent heterogeneities, *Panel B* of Table 5 reports our LNK estimates $\widehat{A}(\cdot)$ at five percentiles of the estimated index $EXS_{it}\widehat{\gamma}_{exs} + \widehat{\alpha}_i$. The function estimates are all significant at 1% level, with the mean given

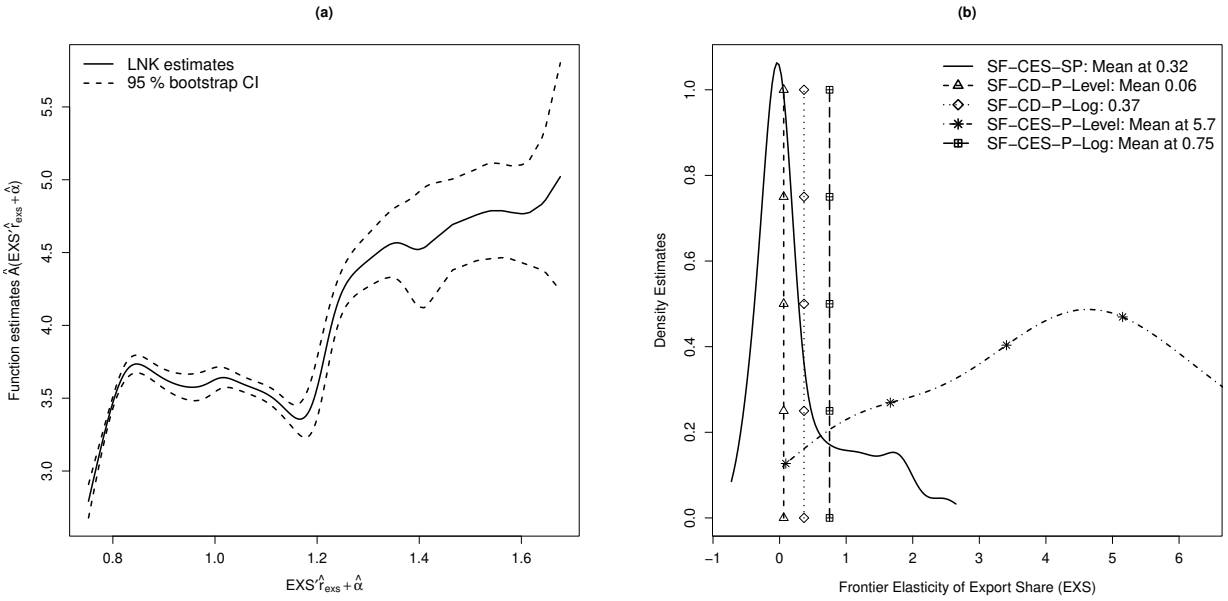
[9]Note that the observation-specific RTS is consistent with the implication from constant RTS estimate $\widehat{\nu}$ in *Panel A* of Table 4. The increasing RTS is suggested under both SF-CES-SP and SF-CES-P-Level, equivalently to the result that $\widehat{\nu}$ is significantly larger than one at 1% level. Similarly, the decreasing RTS in SF-CES-P-Log is in line with the result that $\widehat{\nu}$ is insignificantly different from one.

Table 5: Summary of Estimation Results on Technology

Panel A	SF-CES-SP		SF-CD-P-Level		SF-CD-P-Log		SF-CES-P-Level		SF-CES-P-Log	
EXS	0.6705***	(0.2314)	0.3671***	(0.1442)	0.0922***	(0.0201)	0.7505***	(0.1955)	0.058***	(0.0214)
Panel B	$\widehat{A}(EXS\widehat{\gamma}_{exs}+\widehat{\alpha})$		$\widehat{A}_p(EXS\widehat{\gamma}_{exs}+\widehat{\alpha})$		$\widehat{A}_p(EXS\widehat{\gamma}_{exs}+\widehat{\alpha})$		$\widehat{A}_p(EXS\widehat{\gamma}_{exs}+\widehat{\alpha})$		$\widehat{A}_p(EXS\widehat{\gamma}_{exs}+\widehat{\alpha})$	
0.10	3.4983***	(0.0395)	2.1057***	(0.0208)	1.9943***	(0.0305)	3.6512***	(0.0208)	4.8434***	(0.0219)
0.25	3.7041***	(0.0444)	2.1837***	(0.0212)	2.0126***	(0.0314)	3.9331***	(0.0224)	4.8665***	(0.0289)
0.50	3.5753***	(0.0536)	2.2855***	(0.0325)	2.0358***	(0.0324)	4.3174***	(0.0344)	4.8956***	(0.0482)
0.75	3.5353***	(0.0632)	2.4752***	(0.0258)	2.077***	(0.0342)	5.0816***	(0.0336)	4.9469***	(0.0348)
0.90	3.4549***	(0.0823)	2.6030***	(0.0514)	2.1034***	(0.0354)	5.6328***	(0.0623)	4.9795***	(0.0790)
Mean	3.6291		2.3436		2.0470		4.5896		4.9162	
Panel C	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Elas(EXS)	0.3239	0.0293	0.0572	0.058	0.3637	0.3637	0.7794	0.7794	0.3693	0.3273

Note: Standard errors are reported in parenthesis beneath the estimated coefficients in *Panel A* and *Panel B*. ***,**,* stands for the statistically significant level at 1%, 5% and 10%, respectively.

Figure 4: Technology Function Estimate (a) and Frontier Elasticity of Export Share (b)



on the bottom of the panel. We observe that technology is not increased monotonically when $EXS_{it}\widehat{\gamma}_{exs}+\widehat{\alpha}_i$ exceeds about its 25% percentile. To better visualize the estimated technology function, panel (a) of Figure 4 plots $\widehat{A}(\cdot)$ (solid line) along with its 95% wild bootstrap confidence interval (dashed line) using 399 repetitions. Our estimated index $EXS_{it}\widehat{\gamma}_{exs}+\widehat{\alpha}_i$ is positive for all observations, given that $\widehat{\gamma}_{exs}$ and 17 out of 19 fixed effects estimates are significantly positive.[10] As indicated by Table 5, an initial increase in index (driven by increasing export share) from 0.752

[10]The fixed effect being dropped as a reference point corresponds to Turkey.

to 0.847 (about 25% percentile) significantly improves the technology from 2.793 toward 3.735, or a 33.7% increase. A further rise in the index toward 1.167 (about 88% percentile), however, steadily declines $\widehat{A}$ until it reaches a local trough at 3.356 (about 11.3% decrease). The technology is improved again with export share plus fixed effects continuously rising, although this upward moving trend is driven by less than 10% country-year observations (mostly from the Netherlands and Belgium with high export shares).

Our SF-CES-SP reveals a significant nonlinear structure of the technology function, which clearly deviates from a commonly assumed exponential form. For comparison purpose, *Panel B* of Table 5 reports estimated exponential technology $\widehat{A}_p(\cdot) = \exp(\cdot)$ from all four benchmark models. As expected, $\widehat{A}_p(\cdot)$ differs from its semiparametric counterpart both in magnitude and rate of change because, in this case, the index $EXS_{it}\widehat{\gamma}_{exs} + \widehat{\alpha}_i$ is assumed *a priori* to rise technology monotonically at an exponential rate. As a result, different implications are drawn for the role of EXS in technology. The SF-CES-SP implies that, for the majority of the estimated index ($< 90\%$ percentiles), the technology is significantly increased only when the export share (plus fixed effects) rises toward a moderate level. Given our sample, it suggests that countries with relatively low EXS, such as Turkey, USA, and Greece, are likely to witness technology improvement if export becomes more intensive. This is in contrast to the implication from all benchmark models, suggesting that the technology can be improved at an increasing rate by rising export share. Compared to benchmark models, therefore, modeling technology nonparametrically is vital to avoid misleading inference on the effect of environment variables due to model misspecification.

Given that EXS significantly influences the frontier (F) through technology, we further compute the frontier elasticity of export share, or Elas(EXS), defined as $\frac{dF/F}{dEXS}$ (i.e., the percentage change in frontier due to one unit change in export share).[11] *Panel C* of Table 5 reports the mean and median of Elas(EXS) under each model, which shows a small variation in SF-CD-P-Level but wide dispersion in SF-CES-SP and SF-CES-P-Level. This can be vividly seen from panel (b) of Figure 4, which plots kernel densities of Elas(EXS) under models in levels against constants of Elas(EXS) under models in logs. Here, we only plot the average of Elas(EXS) under SF-CD-P-Level because the estimates are highly centered around the mean. The estimates of Elas(EXS) across all models (except for SF-CES-P-Level) are less than unity with the lowest mean of 0.06 in SF-CD-P-Level, followed by 0.32 in SF-CES-SP, 0.37 in SF-CD-P-Log and 0.75 in SF-CES-P-Log. Different from benchmark models, our SF-CES-SP yields 35.6% negative estimates due to the downward moving trend of $\widehat{A}(\cdot)$ depicted in panel (a) of Figure 3. Different from the rest of the models, the Elas(EXS) under SF-CES-P-Level are exceptionally large with a mean of 5.7 and a notable variation. Overall,

[11] Earlier we defined the frontier as the production function without inefficiency. That is, we have $F = A(EXS\gamma_{exs} + \alpha)\left[\pi K^{-\rho} + (1-\pi)HL^{-\rho}\right]^{-\frac{\nu}{\rho}}$ under SF-CES-SP; $F = A_p(EXS\gamma_{exs} + \alpha)K^{\beta_K}HL^{\beta_{HL}}$ under SF-CD-P-Level; and $F = A_p(EXS\gamma_{exs} + \alpha)\left[\pi K^{-\rho} + (1-\pi)HL^{-\rho}\right]^{-\frac{\nu}{\rho}}$ under SF-CES-P-Level. Notice that under SF-CD-P-Log and SF-CES-P-Log, $\frac{dF/F}{dEXS} = d\ln(F)/dEXS = \gamma_{exs}$. Thus, Elas($EXS$) are observation-specific under three models in level but constant under two models in logs.

Table 6: Summary of Estimation Results on Inefficiency

	SF-CES-SP		SF-CD-P-Level		SF-CD-P-Log		SF-CES-P-Level		SF-CES-P-Log	
Cons	0.5478***	(0.1232)	0.5094***	(0.1432)	0.9856***	(0.1697)	0.1615*	(0.0932)	1.094***	(0.0924)
EXS	-0.2949***	(0.0985)	-0.376***	(0.0623)	-0.0377	(0.0535)	-0.2055**	(0.0853)	-0.0501	(0.5155)
FDI-INS	-0.1052***	(0.0524)	-0.4273***	(0.0448)	-0.1137*	(0.0640)	-0.1138***	(0.0417)	-0.1004*	(0.0532)
FDI-OUTS	0.1276***	(0.0677)	-0.4893***	(0.0572)	-0.0534	(0.0587)	0.1339**	(0.0604)	-0.0482	(0.0312)

Note: Standard errors are reported in parenthesis beneath the estimated coefficients in *Panel A* and *Panel B*. ***,**,* stands for the statistically significant level at 1%, 5% and 10%, respectively.

we observe a significant positive effect of export share on improving country's production frontier.

We finally compare estimation results regarding the inefficiency function. *Panel A* of Table 6 reports the coefficient estimates (i.e., $\widehat{\theta}$) under each model. Recall that $\mu(\cdot) = \exp(\cdot)$ is positive, a coefficient with negative sign indicates that a rise in the corresponding inefficiency determinant reduces the mean inefficiency. In fact, the exponential structure of $\mu(\cdot)$ implies that each $\widehat{\theta}_j$ is the estimated inefficiency elasticity of its (percentage measured) determinant W_j, because $\frac{d\mu(W;\theta)/\mu(W;\theta)}{dW_j} = \theta_j$ for $j \in \{EXS, FDI\text{-}INS, FDI\text{-}OUTS\}$. Our SF-CES-SP shows that, all else constant, the marginal impact of export share on reducing a country's mean inefficiency (or shrinks the averaged gap between a country's frontier and actual output) is 29.5% at the 1% significance level. Combined with our above results on the frontier elasticity of EXS, our model reveals that a higher (yet not excessive) export share is beneficial for not only improving a country's frontier but facilitating the country to produce closer to its frontier. The inefficiency elasticity of EXS remains qualitatively similar in SF-CD-P-Level (-0.376) and SF-CES-P-Level (-0.206), but becomes indifferent from zero once log transformation is applied in SF-CD-P-Log and SF-CES-P-Log.

Similar to the findings in Iyer et al. (2008), we observe opposite roles of FDI inflows and outflow shares in determining the mean inefficiency. On one hand, a one percentage increase in FDI inflows under SF-CES-SP shrinks mean inefficiency by 10.5%, quantitatively similar to all benchmark models but SF-CD-P-Level. On the other hand, a one unit (or fraction) increase in $FDI\text{-}OUTS$ under SF-CES-SP significantly enlarges the mean inefficiency (or widens the averaged gap between a country's frontier and actual output) by 12.8%, fairly close to 13.4% under SF-CES-P-Level. One possible underlying cause for the effect of $FDI\text{-}OUTS$ is that capital outflow may induce producers to opt for low-cost facilities overseas, which deteriorates a country's manufacturing sector, depletes the capital market, and thus lowers the country's efficiency (Iyer et al., 2008). Similar to the effect of EXS in $\mu(\cdot)$, the impact of $FDI\text{-}OUTS$ turns to be either amplified under SF-CD-P-Level or insignificant in log-transformed benchmark models.

The estimated inefficiency function facilitates our analysis on the country-specific productive efficiency. We evaluate the efficiency by ranking countries based on the yearly average of the inefficiency function estimates as $\bar{\widehat{\mu}}(W_i;\widehat{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mu}(W_{it};\widehat{\theta})$. Thus, a more efficient country is

Table 7: Country Ranking based on SF-CES-SP Yearly Averaged Inefficiency Function Estimates

Ranking	SF-CES-SP	$\bar{\widehat{\mu}}(\cdot)$	$\%\Delta\bar{\widehat{\mu}}(\cdot)$	SF-CD-P-Level	SF-CD-P-Log	SF-CES-P-Level	SF-CES-P-Log
1	USA	1.3261	-0.04%	USA	USA	Luxembourg	Luxembourg
2	Luxembourg	1.4397	0.02%	Luxembourg	Luxembourg	USA	USA
3	Belgium	1.4469	-0.06%	Belgium	Belgium	Belgium	Belgium
4	Netherlands	1.4784	-0.35%	Netherlands	Ireland	Netherlands	Ireland
5	Ireland	1.5153	-0.21%	Ireland	Netherlands	Ireland	Netherlands
6	Norway	1.5174	0.09%	Norway	Norway	Norway	Norway
7	Denmark	1.5282	-0.18%	Switzerland	Switzerland	Denmark	Switzerland
8	Sweden	1.5332	-0.11%	Denmark	Sweden	Sweden	Denmark
9	Switzerland	1.5430	-0.30%	Sweden	Denmark	Switzerland	Sweden
10	Iceland	1.5453	-0.08%	Iceland	Iceland	Iceland	Iceland
11	Austria	1.5514	-0.22%	Austria	Austria	Austria	Austria
12	Germany	1.5712	-0.26%	Germany	Germany	Germany	Germany
13	Canada	1.6144	-0.06%	Canada	Canada	Canada	Canada
14	France	1.6154	-0.10%	UK	UK	France	UK
15	UK	1.6211	-0.06%	France	France	UK	France
16	Italy	1.6328	-0.11%	Italy	Italy	Italy	Italy
17	Portugal	1.6496	-0.14%	Portugal	Portugal	Portugal	Portugal
18	Spain	1.6773	-0.14%	Spain	Spain	Spain	Spain
19	Greece	1.6863	-0.10%	Greece	Greece	Greece	Greece
20	Turkey	1.6927	-0.46%	Turkey	Turkey	Turkey	Turkey

indicated by a lower magnitude of $\bar{\widehat{\mu}}(\cdot)$.[12] We also compute the yearly average of percentage change in inefficiency function estimates as $\%\Delta\bar{\widehat{\mu}}(W_i;\widehat{\theta}) = \frac{1}{T-1}\sum_{t=2}^{T}(\widehat{\mu}(W_{it};\widehat{\theta}) - \widehat{\mu}(W_{it-1};\widehat{\theta}))/\widehat{\mu}(W_{it-1};\widehat{\theta})$. The first four columns in Table 7 report the rank of OECD countries based on an ascending order of $\bar{\widehat{\mu}}(\cdot)$ under SF-CES-SP, along with the corresponding $\%\Delta\bar{\widehat{\mu}}(\cdot)$. The USA is ranked the most efficient country over the past five decades, in line with observations in Henderson and Russell (2005) that USA produces on the frontier in 1965 and 1990. Luxembourg, Belgium, and the Netherlands are also fairly efficient, but the Netherlands discloses a faster decreasing rate of inefficiency of $\%\Delta\bar{\widehat{\mu}}(\cdot) = -0.35\%$. Spain and Greece are more inefficient, placed at the third and second to last of the ranking, respectively. Although Turkey is ranked the most inefficient country, it catches up the frontier (or decreases inefficiency) at the fastest average rate of 0.46% compared to all other OECD countries. As reported in the last four columns of Table 7, we do not observe significant differences in the ranking across all four benchmark models, particular for the top and bottom three countries.

All comparisons above disclose notable estimation differences across SF-CES-SP and benchmark models. The differences may be driven by two sources of estimation bias in benchmark models. The first source is the log approximation on a SF model, which impedes the identification of parameters in the frontier and inefficiency function (see (9) and (11)). The log-induced bias is larger in SF-CES-P-Log because, as shown in (11), the log-approximation error stems from both frontier and composite error that are not asymptotically negligible. This may explain our findings

[12] See **Remark 3** for alternative measures for technical efficiency in our model.

that SF-CES-P-Log differs significantly from all other models in terms of frontier and inefficiency estimation. The second source is the exponential function imposed *a priori* on technology, which masks an important non-monotonic effect (and thus implication) of export share as revealed by our SF-CES-SP. To provide statistical evidence on the potential misspecification of $A_p(\cdot) = \exp(\cdot)$, we implement a nonparametric test for the null hypothesis $H_0 : Pr(A(v) = \exp(v)) = 1$ for almost all $v = EXS\gamma_{exs} + \alpha$. Following the insights of Ullah (1985), we construct the F-type test statistic $J = (SSR_0 - SSR_1)/SSR_1$, where SSR_0 is the sum of squared residuals (SSR) of SF-CES-P-Level under H_0 and and SSR_1 the SSR of SF-CES-SP under the alternative.[13] Notice that SF-CES-SP properly nests SF-CES-P-Level as a special case. We employ a bootstrap version of J with 399 repetitions, and obtain $J = 64.952$ with empirical p-value of 0.0000. Thus, we reject the null of a correctly specified $A_p(\cdot) = \exp(\cdot)$ at 1% level, indicating that an exponential function is inadequate to capture the evolvement of technology. Overall, our empirical findings highlight the importance of considering a semiparametric CES frontier in level, as the case in SF-CES-SP, for productive efficiency analysis.

5 Conclusion

We propose a panel SF model (SF-CES-SP) to alleviate parametric restrictions on both the frontier and composite error. Our frontier specifies a semiparametric form of the CES production function, which improves the conventional CES by avoiding log-induced estimation bias and allowing unknown functional form of technology. The semiparametric CES frontier specifies the technology as a smooth single-index function of environment variables, and within the index, the fixed effects are accounted for individual unobserved, time invariant heterogeneity. The frontier further retains certain parameters of interests to be directly estimated, including distribution of inputs, elasticity of inputs substitution, and returns to scale. Our composite error requires no distributional assumptions on either the stochastic noise or inefficiency, and allows the conditional mean of the inefficiency to depend on a wide range of determinants. The structure of our proposed frontier model is flexible in modeling the complex nonlinear structure of a production function, free of the curse of dimensionality, and robust to different distributions of inefficiency. The appealing finite-sample performance of the proposed estimator is demonstrated through simulation studies.

We showcase the empirical applicability of SF-CES-SP by investigating productive efficiency among 20 OECD countries from 1970-2019. Our model provides reasonable estimation on several metrics of the frontier, suggesting an increasing returns to scale production function, and complementary relationship between capital and (human capital augmented) labor. The flexible structure

[13] We construct the SSR under H_0 as $SSR_0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[Y_{it} - \widehat{F}_{it} + \mu(W_{it};\widehat{\theta})\right]^2$, where $\widehat{F}_{it} = A_p(EXS_{it}\widehat{\gamma}_{exs} + \widehat{\alpha}_i)\left[K_{it}^{-\widehat{\rho}} + HL_{it}^{-\widehat{\rho}}\right]^{-\frac{\widehat{\nu}}{\widehat{\rho}}}$ is the estimated CES frontier in SF-CES-P-Level. The construction of SSR_1 is similar to SSR_0, except that $\widehat{F}_{it}$ is replaced with estimated frontier from SF-CES-SP assuming unknown structure of $A(\cdot)$.

of our model uncovers a significantly nonlinear technology function. We find that moderate increase in export share significantly rises a country's frontier and reduces productive inefficiency. The inefficiency can be also reduced by a higher FDI inflow and lower FDI outflow share. Compared to our proposed model, we observe large different (and perhaps unreasonable in some cases) estimation from conventional parametric SF models. Through results comparison, we show that a SF model with semiparametric CES frontier in level is more robust to estimation bias due to model misspecification and log-transformation.

References

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

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts.

Arellano, M., Hahn, J. 2007. Understanding bias in nonlinear panel models: Some recent developments. Econometric Society Monographs, 43, 381.

Baltagi, B. H., Li, D., 2002. Series estimation of partially linear panel data models with fixed-effects. Annals of Economics and Finance, 3, 103–116.

Battese, G. E., Coelli, T. J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics, 38 (3), 387–399.

Battese, G. E., Coelli, T. J., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 3, 153–169.

Cornwell, C., Schmidt, P., Sickles, R. C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46 (2), 185–200.

de Boor, C. R., 1978. *A Practical Guide to Splines*. New York: Springer-Verlag.

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business and Economic Statistics 14, 460–468.

Ferrara, G., Vidoli, F., 2017. Semiparametric stochastic frontier models: A generalized additive model approach. European Journal of Operational Research 258 (2), 761–777.

Fernández-Val, I., Weidner, M., 2016. Individual and time effects in nonlinear panel models with large N, T. Journal of Econometrics, 192 (1), 291–312.

Gollin, D., 2002. Getting income shares right. Journal of Political Economy, 110, 458–474.

Henderson, D. J., 2009. A non-parametric examination of capital–skill complementarity. Oxford Bulletin of Economics and Statistics, 71 (4), 519–538.

Henderson, D. J., Russell, R. R., 2005. Human capital and convergence: A production-frontier approach. International Economic Review, 46 (4), 1167–1205.

Henderson, D. J., Parmeter, C. F., 2015. *Applied Nonparametric Econometrics*. New York: Cambridge University Press.

Horrace, W. C., Parmeter, C. F., 2011. Semiparametric deconvolution with unknown error variance. Journal of Productivity Analysis 35 (2), 129–141.

Iyer, K. G., Rambaldi, A. N., Tang, K. K., 2008. Efficiency externalities of trade and alternative forms of foreign investment in OECD countries. Journal of Applied Econometrics 23 (6), 749–766.

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

Karabarbounis, L., Neiman, B., 2013. The global decline of the labor share. The Quarterly Journal of Economics 129 (1), 61–103.

Kneip, A., Simar, L., 1996. A general framework for frontier estimation with panel data. Journal of Productivity Analysis, 7 (2), 187–212.

Kneip, A., Simar, L., Van Keilegom, I., 2015. Frontier estimation in the presence of measurement error with unknown variance. Journal of Econometrics 184 (2), 379–393.

Kumbhakar, S. C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46 (1-2), 201–211.

Kumbhakar, S. C., Lovell, C. A. K., 2000. Stochastic frontier analysis. Cambridge university Press, Cambridge, UK.

Kumbhakar, S. C., Wang, H.-J., 2005a. Estimation of growth convergence using a stochastic production frontier approach. Economics Letters 88 (3), 300–305.

Kumbhakar, S. C., Park, B. U., Simar, L., Tsionas, E., 2007. Nonparametric stochastic frontiers: a local maximum likelihood approach. Journal of Econometrics 137, 1–27.

Kumbhakar, S. C., Wang, H. J., Horncastle, A. P., 2015. A Practitioner's Guide to Stochastic Frontier Analysis Using STATA. Cambridge University Press.

Kumbhakar, S. C., Parmeter, C. F., Zelenyuk, V., 2020. Stochastic frontier analysis: Foundations and advances I-II. Handbook of Production Economics, 1–40.

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

Lee, Y., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical efficiency. In: in L. H. Fried & S. Schmidt, eds, 'The Measurement of Productive Efficiency'. Oxford University Press, Oxford, United Kingdom.

Ma, S., Song, P. X., 2015. Varying index coefficient models. Journal of the American Statistical Association, 110, 341–356.

Malikov, E., Zhao, S., Kumbhakar, S. C., 2020. Estimation of firm-level productivity in the presence of exports: Evidence from China's manufacturing. Journal of Applied Econometrics, 35 (4), 457–480.

Martins-Filho, C., Yao, F., 2015. Nonparametric stochastic frontier estimation via profile likelihood. Econometric Reviews 34 (4), 413–451.

Masanjala, W. H., Papageorgiou, C., 2004. The Solow model with CES technology: Nonlinearities and parameter heterogeneity. Journal of Applied Econometrics, 19 (2), 171–201.

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

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

Park, B. U., Simar, L., Zelenyuk, V., 2015. Categorical data in local maximum likelihood: Theory and applications to productivity analysis. Journal of Productivity Analysis 43, 199–214.

Parmeter, C. F., Kumbhakar, S. C., 2014. Efficiency Analysis: A Primer on Recent Advances. Foundations and Trends® in Econometrics 7 (3-4), 191–385.

Parmeter, C. F., Wang, H.-J., Kumbhakar, S. C., 2016. Nonparametric estimation of the determinants of inefficiency. Journal of Productivity Analysis.

Santos Silva, J. M. C., Tenreyro, S. 2006. The log of gravity. Review of Economics and Statistics, 88, 641–658.

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

Sun, K., Henderson, D. J., Kumbhakar, S., 2008. Biases in approximating log production. Journal of Applied Econometrics, 26, 708–714.

Sun, K., Kumbhakar, S. C., 2013. Semiparametric smooth-coefficient stochastic frontier model. Economics Letters 120 (2), 305–309.

Tran, K. C., Tsionas, E. G., 2009. Estimation of nonparametric inefficiency effects stochastic frontier models with an application to British manufacturing. Economic Modelling, 26, 904—909.

Ullah, A. 1985. Specification analysis of econometric models. Journal of Quantitative Economics 1 (2), 187–209.

Wang, T., Tian, J., Yao, F., 2020. Does high leverage ratio influence Chinese firm performance? A semiparametric stochastic frontier approach with zero-inefficiency. Empirical Economics, 1–50.

Wang, T., Henderson, D. J., 2022. Estimation of a varying coefficient, fixed-effects Cobb-Douglas production function in levels. Economics Letters, 1–6.

Yao, F., Zhang, F., Kumbhakar, S. C., 2018a. Semiparametric smooth coefficient stochastic frontier model with panel data. Journal of Business & Economic Statistics, 1–17.

Yao, F., Wang, T., Tian, J., Kumbhakar, S. C., 2018b. Estimation of a smooth coefficient zero-inefficiency panel stochastic frontier model: A semiparametric approach. Economics Letters 166, 25–30.

Zhu, L., Xue, L., 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68, 549–570.

Zhou, J., Parmeter, C. F., Kumbhakar, S. C. 2020. Nonparametric estimation of the determinants of inefficiency in the presence of firm heterogeneity. European Journal of Operational Research, 286 (3), 1142-1152.

Appendix A: Incidental Parameters Problem

From our SF-CES-SP in (1), the estimation of individual fixed effects α_i, appearing in $A(\cdot)$, can be challenging with large n and fixed T (i.e., the incidental parameters problem). We expect worse numerical performance of our proposed estimator as α_i is not consistently estimated. To investigate the impact of incidental parameters problem on our model estimation, we repeat the same exercise for our simulation studies in Section 3, except that we fix $T = 10$ and let $n = (10, 20, 40)$.

Table A1 reports the simulation results, which shows a clear negative impact of the incidental parameters problem on the estimator's performance. We note that a large $n = 40$ with a small $T = 10$ leads to a much worse estimation of the fixed effects, causing parameter estimates, such as $\widehat{\gamma}_{z_1}$, fail to converge. For instance, the RMSE of $\widehat{\gamma}_{z_1}$ under DGP_1 with $(n, T) = (40, 10)$ increases by more than five fold compared to that with $(n, T) = (10, 40)$ in Table 1. This observation is consistent with conventional result in nonlinear parametric panel models, where the incidental parameters impact not only fixed effects but other coefficients when the fixed effects enter the model nonlinearity (Arellano and Hahn, 2007; Fernández-Val and Vidoli, 2016). Since the LNK estimator $\widehat{A}(\cdot)$ depends crucially on the accurate measure of the index $Z_{it}^{\top}\gamma_z + \alpha_i$, the approximation error on technology is evidently inflated as n doubles across all DGPs due to the poor estimation of α_i.

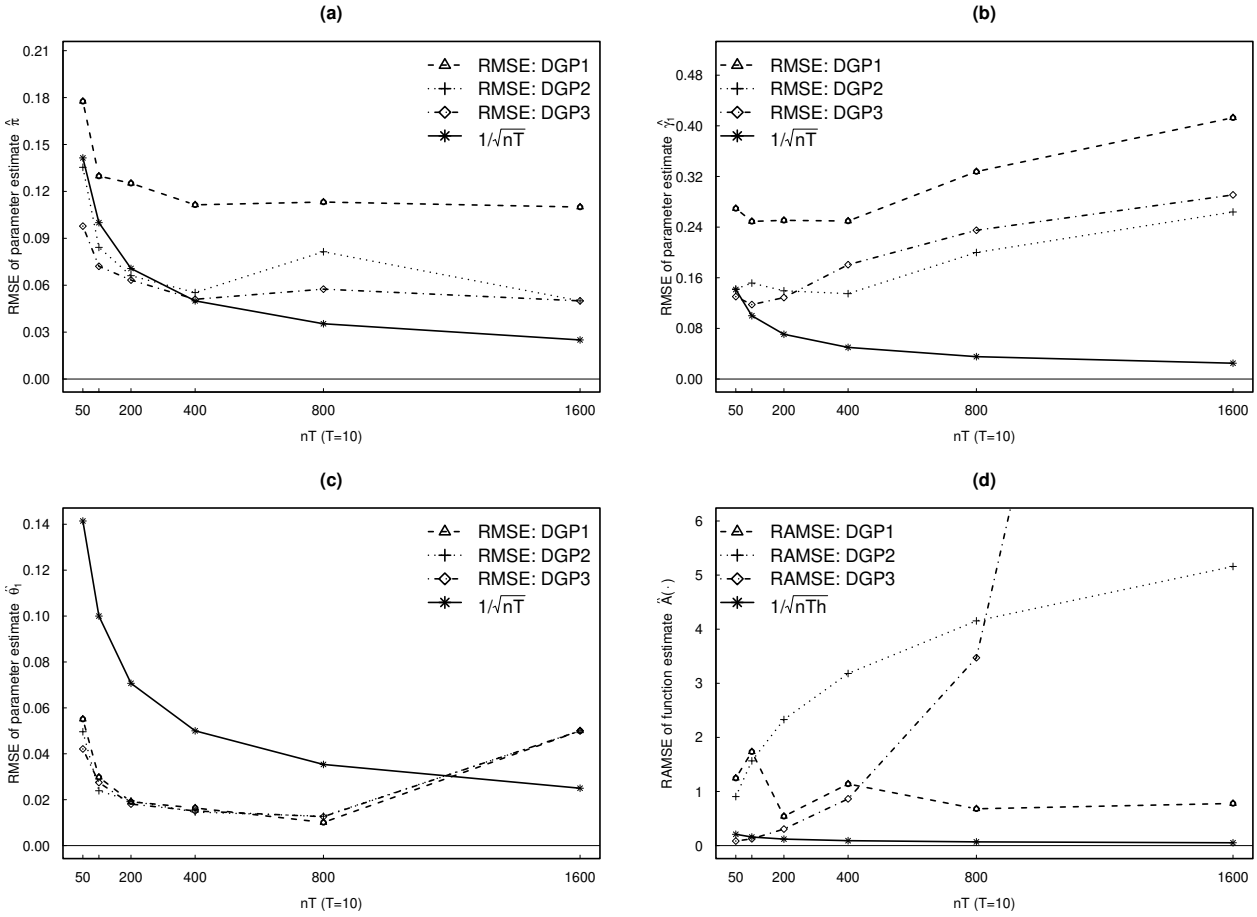
As a consequence, the model estimator becomes inconsistent due to the presence of incidental parameters. We demonstrate it by repeating our simulations with $T = 10$ and $n = (5, 10, 20, 40, 80, 160)$, and compute RMSE (RAMSE) of parameter (function) estimates. Similar to Figure 1, we plot in Figure A1 the RMSE of $\widehat{\pi}$, $\widehat{\gamma}_{z_1}$, and $\widehat{\theta}_1$ against the parametric rate in panel (a), (b), and (c), respectively, under DGP_{1-3}. Clearly, each parameter estimator no longer converges at a rate of $1/\sqrt{nT}$ as n doubles, regardless of which DGP being considered. Panel (d) of Figure A1 plots the RAMSE of function estimates against the nonparametric rate under each DGP. In constrast to our observation in Figure 1, the LNK estimator is now largely biased without converging at a rate of $1/\sqrt{nTh}$.

Overall, our results should be taken as numerical evidence for applied researchers to avoid having a large n with small T for a panel data model with fixed-effects (i.e., a large ratio of n/T) in order to achieve reasonable estimation for our SP-CES-SP.

Table A1: Incidental Parameters Problem in SF-CES-SP with Scaling Inefficiency in DGP_{1-3}

	(n,T)		RMSE	BIAS	STD		RAMSE	ABIAS	ASTD
DGP_1	$(10,10)$	$\widehat{\gamma}_{z_1}$	0.2442	0.1931	0.1502	$\widehat{A}$	2.5605	1.4440	2.5348
		$\widehat{\pi}$	0.1610	0.1299	0.0955				
		$\widehat{\rho}$	0.1012	0.0922	0.0421				
		$\widehat{\nu}$	0.0872	0.0750	0.0449				
		$\widehat{\theta}_1$	0.0946	0.0750	0.0580				
	$(10,20)$	$\widehat{\gamma}_{z_1}$	0.2355	0.1833	0.1486	$\widehat{A}$	3.3007	1.1071	3.1073
		$\widehat{\pi}$	0.1207	0.0978	0.0712				
		$\widehat{\rho}$	0.1139	0.0919	0.0676				
		$\widehat{\nu}$	0.0774	0.0705	0.0322				
		$\widehat{\theta}_1$	0.0736	0.0625	0.0390				
	$(10,40)$	$\widehat{\gamma}_{z_1}$	0.2799	0.2345	0.1536	$\widehat{A}$	1.3053	0.1829	0.6652
		$\widehat{\pi}$	0.1106	0.0831	0.0734				
		$\widehat{\rho}$	0.0851	0.0820	0.0228				
		$\widehat{\nu}$	0.0702	0.0672	0.0206				
		$\widehat{\theta}_1$	0.0660	0.0582	0.0313				
DGP_2	$(10,10)$	$\widehat{\gamma}_{z_1}$	0.1275	0.0914	0.0893	$\widehat{A}$	4.5039	1.8774	4.3394
		$\widehat{\pi}$	0.0971	0.0788	0.0570				
		$\widehat{\rho}$	0.1004	0.0825	0.0575				
		$\widehat{\nu}$	0.0813	0.0656	0.0483				
		$\widehat{\theta}_1$	0.0788	0.0656	0.0439				
	$(10,20)$	$\widehat{\gamma}_{z_1}$	0.1383	0.0963	0.0997	$\widehat{A}$	14.8140	2.7401	14.4186
		$\widehat{\pi}$	0.0696	0.0538	0.0444				
		$\widehat{\rho}$	0.1069	0.0935	0.0521				
		$\widehat{\nu}$	0.0613	0.0477	0.0387				
		$\widehat{\theta}_1$	0.0645	0.0530	0.0369				
	$(10,40)$	$\widehat{\gamma}_{z_1}$	0.1591	0.1133	0.1123	$\widehat{A}$	2.3722	2.3582	1.9548
		$\widehat{\pi}$	0.0536	0.0427	0.0327				
		$\widehat{\rho}$	0.0905	0.0852	0.0304				
		$\widehat{\nu}$	0.0585	0.0501	0.0305				
		$\widehat{\theta}_1$	0.0599	0.0510	0.0317				
DGP_3	$(10,10)$	$\widehat{\gamma}_{z_1}$	0.1023	0.0730	0.0721	$\widehat{A}$	2.2381	1.2774	2.2072
		$\widehat{\pi}$	0.0831	0.0664	0.0501				
		$\widehat{\rho}$	0.1570	0.1216	0.0997				
		$\widehat{\nu}$	0.0962	0.0720	0.0642				
		$\widehat{\theta}_1$	0.0855	0.0719	0.0466				
	$(10,20)$	$\widehat{\gamma}_{z_1}$	0.1301	0.1013	0.0820	$\widehat{A}$	5.8606	2.0761	5.7200
		$\widehat{\pi}$	0.0636	0.0497	0.0399				
		$\widehat{\rho}$	0.1035	0.0887	0.0536				
		$\widehat{\nu}$	0.0527	0.0443	0.0286				
		$\widehat{\theta}_1$	0.0653	0.0553	0.0349				
	$(10,40)$	$\widehat{\gamma}_{z_1}$	0.1939	0.1686	0.0963	$\widehat{A}$	1.0532	0.9806	1.0077
		$\widehat{\pi}$	0.0511	0.0400	0.0320				
		$\widehat{\rho}$	0.0907	0.0811	0.0408				
		$\widehat{\nu}$	0.0555	0.0486	0.0269				
		$\widehat{\theta}_1$	0.0621	0.0532	0.0321				

Figure A1: Incidental parameters problem in Convergence Rates of Estimators in DGP_{1-3}



Appendix B: Demonstration of our R Package

The proposed estimation procedure for SF-CES-SP in (1) is implemented in an **R** package with user-friendly design. The package is available at https://694160821.wixsite.com/taining, and navigate to the top panel of **R code**. The code is stored in a text file and needs to be copy-pasted into an R script for it to run. The file contains five packages (on the top of the file) that need to be pre-installed before running the code for SF-CES-SP estimation. Below these packages, there are a total of 12 functions defined: Function [8.1] generates simulated panel data for demonstration purpose; function [8.2] generates simulated pool data; function [9], which uses functions [1]-[7], performs estimation for SF-CES-SP; and function [10] summarizes the estimation result by a table and several figures.

As an example, we use function [8.1] to generate a panel dataset according to a simple DGP as

$$Y_{it} = A(Z_{it}^{\top}\gamma_z + \alpha_i)\left(\pi X_{1it}^{-\rho} + (1-\pi)X_{2it}^{-\rho}\right)^{-\frac{\nu}{\rho}} + v_{it} - u_{it}, \tag{.1}$$

where $Z_{it} = (Z_{1it}, Z_{2it})$ and X_{jit} are drawn independently from $U(1,4)$. The parameters in the coefficient functions are $\gamma = (2, 4, \alpha_{-1}^{\top})$, with $\alpha_{-1} = (\alpha_2, ..., \alpha_n)^{\top}$. We empirically rescale γ to satisfy $||\gamma_0|| = 1$. We specify $\mu(W_{it};\theta) = \exp(W_{it}^{\top}\theta)$, where $W_{it} = (Z_{1it}, Q_{it})$ with $Q_{it} \sim U(1,4)$. We generate the fixed effects as $\alpha_i = \frac{1}{T}\sum_{t=1}^{T}(X_{1it} + Z_{1it}) + \xi_i$, where $\xi_i \sim \mathcal{N}(0,1)$. Finally, $v_{it} \sim \mathcal{N}(0, 0.5^2)$ is drawn independently from all X, Z, and Q. We specify the technology function $A(v) = \sqrt{v\pi}$. To reduce computational cost in this demonstration, we simply set $(n, T) = (5, 50)$ and generate data as

```
> set.seed(03)
> n=5; t=50
> Dgp<-psudo.data.panel(n=n,t=t,a=1,b=4,c=1,d=4,u.dist=1)
> y<-Dgp$y; x<-Dgp$x; z<-Dgp$z[,c(1:2)]; w<-Dgp$w
```

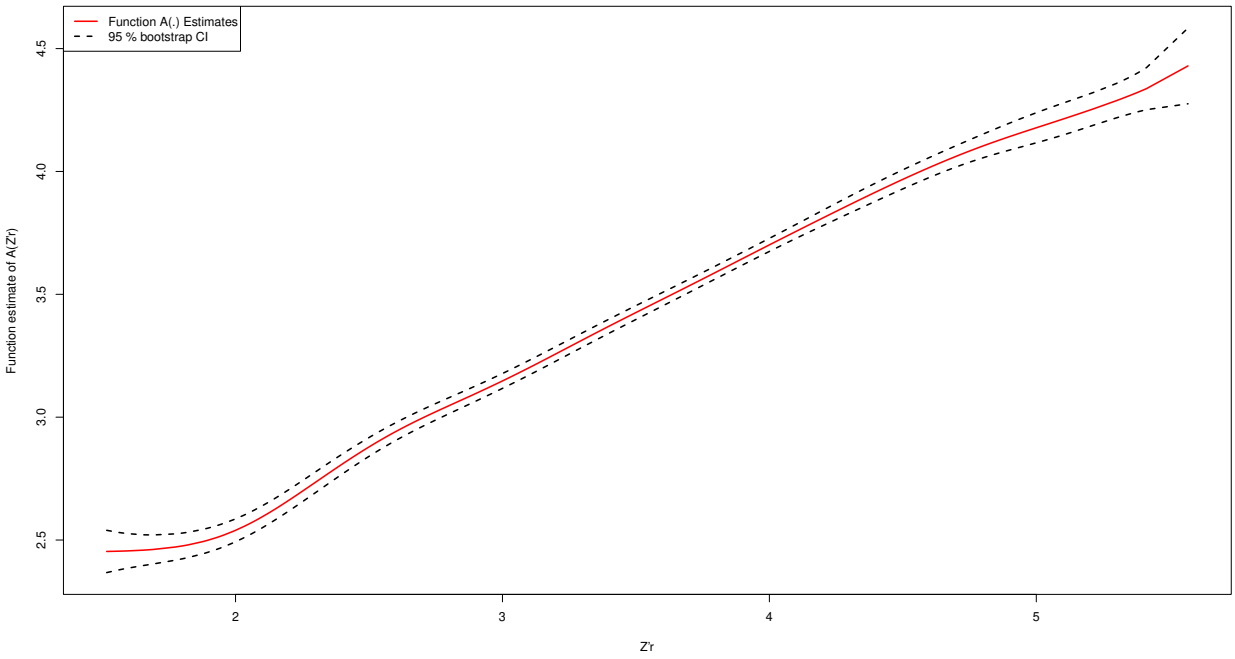
In the first line, a random seed is specified for results replication; in the second line, n and t represent the number of individual and time units, respectively; in the third line, (a,b) are the lower and upper bound of a uniform distribution (i.e., $\mathcal{U}(a,b)$) for each X_j. Similarly, (c,d) are the lower and upper bound of a uniform distribution (i.e., $\mathcal{U}(c,d)$) for all Z_j and Q; in the fourth line, we extract data from Dgp, including a $nT \times 1$ vector y, a $nT \times 2$ matrix x; a $nT \times 2$ matrix z; and a $nT \times 2$ matrix w. The specification u.dist=1 assumes a scaling function of u_{it}, and u.dist=2 sets u_{it} to be Log-Probit (see our specification in Section 3). Running the following code performs estimation for SF-CES-SP in (.1):

```
> sf.ces.sp.panel<-WH.SF.CES(y=y,x=x,z=z,w=w,n=n,t=t,Kernel=TRUE,
Jn.zr.pos = 'quantile',method='fe',u.dist='scaling')

> country.id<-paste0('country',c(1:n))
> result.WH.SF.CES(sf.ces.sp.panel,country.id = country.id)
```

where WH.SF.CES() estimates the SF-CES-SP and saves it as sf.ces.sp.panel, and result.WH.SF.CES() reports the summary of empirical results. Namely, WH.SF.CES() is implemented using data from Dgp above, with 99 wild bootstraps for 95% point-wise confidence interval for LNK estimator. The default number of bootstrap repetitions is boot.num = 99. We purposely set a small boot.num in this example to increase computational speed, but it should be a large integer in practice, such as 399. Notice that WH.SF.CES() reports LNK estimation if Kernel=TRUE, but it also reports series estimation on $A(\cdot)$ if Kernel=FALSE. Therefore, given a large sample data in practice, one can take

Figure A2: Kernel Estimation of $A(\cdot)$ in the Package

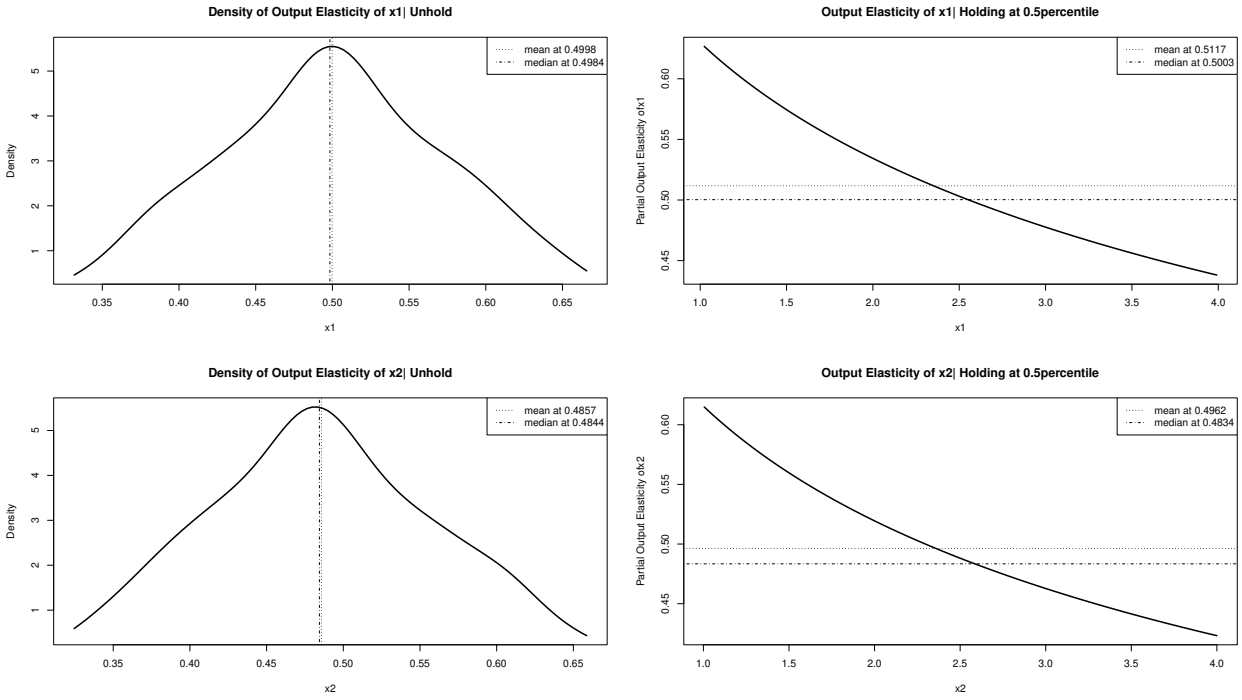


advantage of series estimator to quickly obtain preliminary results without spending much longer time for the LNK estimator. Within WH.SF.CES(), the command Jn.zr.pos = 'quantile' requires the interior knots to be placed on different percentiles of the index $Z_{it}^{\top}\gamma_z + \alpha_i$. If method='fe', a total of $n-1$ dummy variables will be added into the matrix z to control for individual fixed effects. Alternatively, one can perform a pool estimation by setting method='pool', n=nT, and t=1. In this way the fixed effects are assumed zero. The exponential structure of $\mu(\cdot) = \exp(\cdot)$ is specified through u.dist='scaling', and the Log-Probit structure of $\mu(\cdot) = -\ln(\Phi(\cdot))$ through u.dist='log-probit'. All additional arguments in WH.SF.CES() are listed with explanation in the script.

Function result.WH.SF.CES() in turn summarizes the estimation results from sf.ces.sp.panel. It automatically returns numerical results as a table with panels A-D, printed in the console panel of R. Panel A summarizes the estimated parameters in the CES function, technology, and inefficiency mean function. Panel B summarizes the distribution of output elasticity of inputs and return to scales. Panel C reports the summary for estimated technology function and its derivative. Below it are distribution summaries of partial effect of environment variables on frontier.[14] Panel D summarizes different metrics on the estimated inefficiency function. On the bottom of Panel D, a ranking (in ascending order) on individual units based on the magnitude of yearly averaged $\widehat{\mu}(\cdot)$ is

[14]Note that the results reported in Panel C are *partial* effects of each environment variable Z_j on the frontier (i.e., dF/dZ_j for $j = 1, ..., d_z$).

Figure A3: Output Elasticity of Inputs in the Package



displayed, provided that the name of the individual units are given in result.WH.SF.CES(). Here, we simply name them as "country 1,...,country 5" for illustration by generating country.id above, and insert it in result.WH.SF.CES(). The very last panel in the table provides data information, tuning parameters, kernel function type, and nonparametric R-square ($corr(Y_{it}, \widehat{Y}_{it})^2$, with $\widehat{Y}_{it}$ being the fitted SF-CES-SP by LNK estimation.)

Function result.WH.SF.CES() also prints out several figures by default (figure=TRUE). There are four sets of figures to be printed. Each set of figures will be displayed one at a time, and the next set is printed by hitting <Enter>. The first figure is given in Figure A2, which plots the estimated technology function (red solid line) against its point-wise 95% wild bootstrap confidence interval (black dash line). The second set of figures plots output elasticity of inputs (of size $2 \times d_x$), as shown in Figure A3. In the current example, the first column shows the kernel density of output elasticity of X_1 (first row) and of X_2 (second row). The mean and median of the elasticity of inputs are labelled as vertical dot or dashed lines, respectively, and reported in the legend. The second column of Figure A3 plots the partial elasticity of inputs by holding all other variables in X at their respective medians. Again, the mean and median are labelled as horizontal lines. The third set of figures is of size $1 \times d_z$, which plots the kernel density of estimated partial effect of Z on frontier as shown in Figure A4. These estimates are computed by holding all other environment variables in Z at median. The last set of figures is of size $2 \times d_w$, which summarizes the partial effect of

Figure A4: Partial Effects of environment Variables on Frontier in the Package

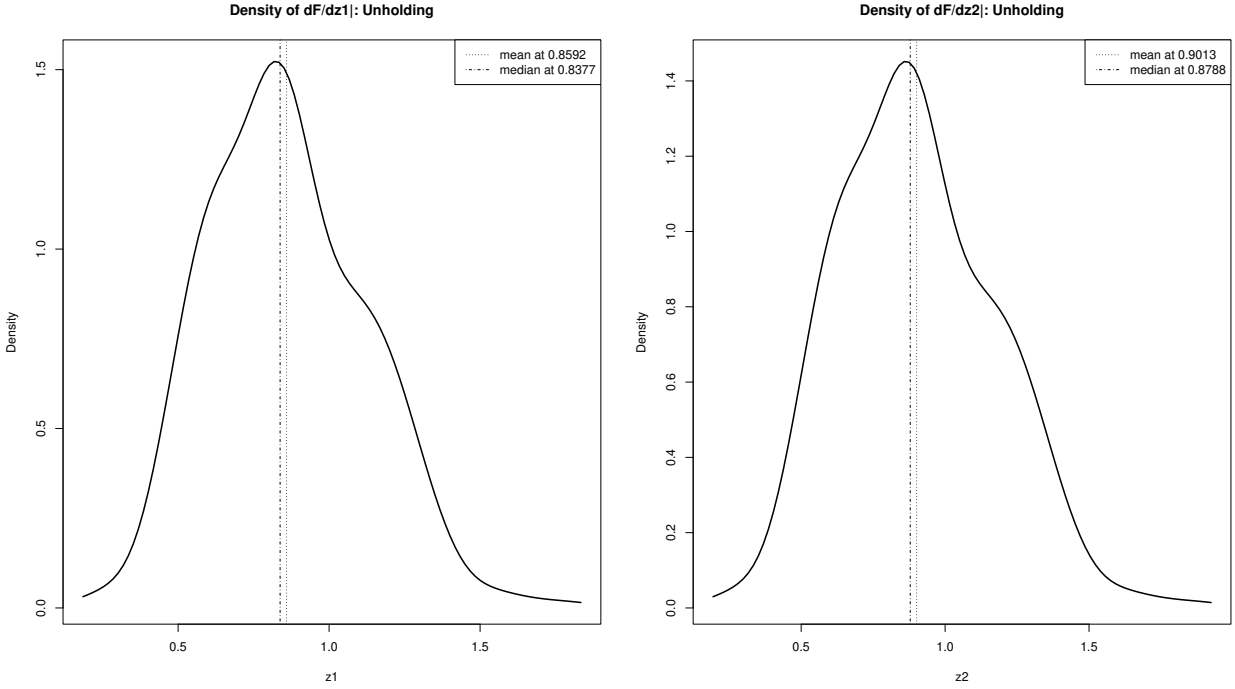
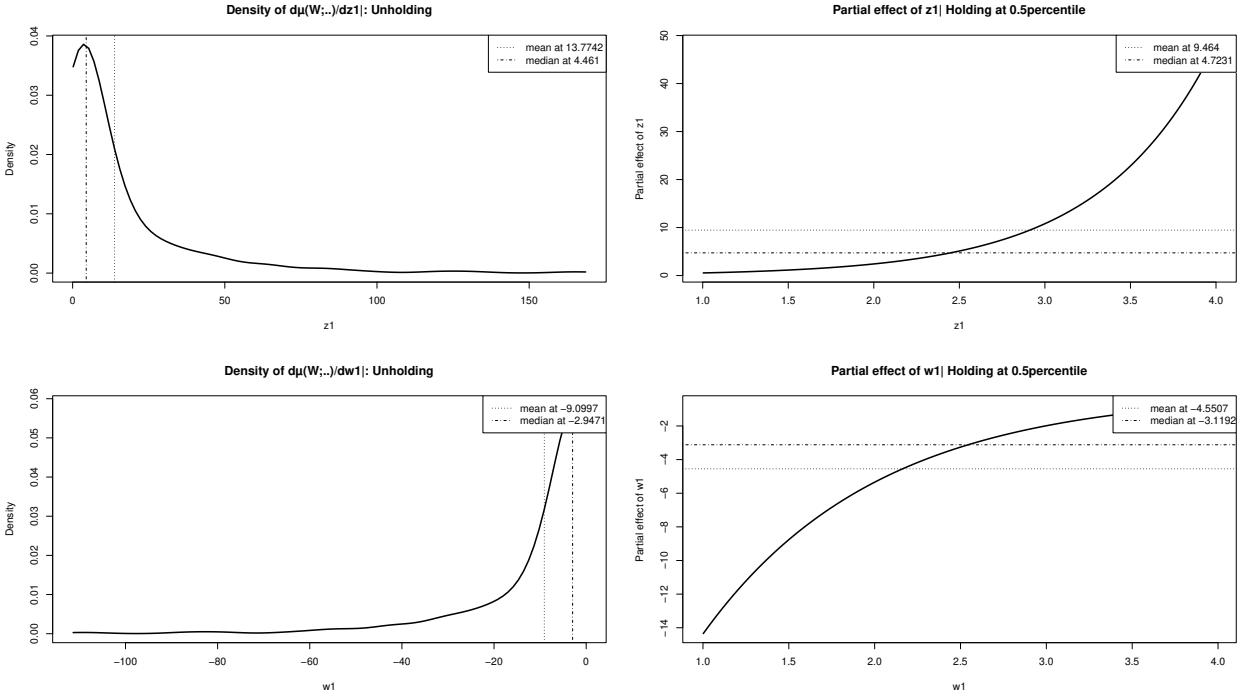


Figure A5: Partial Effects of Inefficiency Determinants in the Package



W on inefficiency mean function as in Figure A5. Similar to Figure A3, the first (second) column reports the kernel density of estimated $d\mu(W_{it};\widehat{\theta})/dW_j$ without (by) holding all else variables in W at certain level (median).

All other arguments in result.WH.SF.CES() are given with explanations in the script. For further questions regarding the implementation of the package, please contact the first author via email.