



# Estimation of semi- and nonparametric stochastic frontier models with endogenous regressors

Artem Prokhorov<sup>1,2</sup> · Kien C. Tran<sup>3</sup> · Mike G. Tsionas<sup>4</sup>

Received: 28 February 2020 / Accepted: 8 September 2020 / Published online: 24 September 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

This paper considers the problem of estimating a nonparametric stochastic frontier model with shape restrictions and when some or all regressors are endogenous. We discuss three estimation strategies based on constructing a likelihood with unknown components. One approach is a three-step constrained semiparametric limited information maximum likelihood, where the first two steps provide local polynomial estimators of the reduced form and frontier equation. This approach imposes the shape restrictions on the frontier equation explicitly. As an alternative, we consider a local limited information maximum likelihood, where we replace the constrained estimation from the first approach with a kernel-based method. This means the shape constraints are satisfied locally by construction. Finally, we consider a smooth-coefficient stochastic frontier model, for which we propose a two-step estimation procedure based on local GMM and MLE. Our Monte Carlo simulations demonstrate attractive finite sample properties of all the proposed estimators. An empirical application to the US banking sector illustrates empirical relevance of these methods.

**Keywords** Constrained semiparametric limited information MLE · Efficiency · Endogeneity · Local limited information MLE · Smooth coefficient · Stochastic frontier

---

✉ Artem Prokhorov  
artem.prokhorov@sydney.edu.au

Kien C. Tran  
kien.tran@uleth.ca

Mike G. Tsionas  
m.tsionas@lancaster.ac.uk

<sup>1</sup> Business School, University of Sydney, Sydney, NSW, Australia

<sup>2</sup> St. Petersburg State University, St. Petersburg, Russia

<sup>3</sup> Department of Economics, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada

<sup>4</sup> Lancaster University Management School, Lancaster LA1 4YX, UK

**JEL Classification** C13 · C14 · C36

## 1 Introduction

In most applications in stochastic frontier model, researchers typically assume that the functional form of the frontier takes a specific parametric form (e.g., Cobb–Douglas, translog, constant elasticity of substitution, etc.) for the data they want to analyze. However, economic theories rarely predict specific functional form for the frontier, and consequently, misspecification of the functional form can lead to biased and inconsistent parameter estimators as well as to misleading inference on inefficiencies. In such cases, researchers may wish to adopt a semi- or nonparametric specification.<sup>1</sup>

Nonparametric stochastic frontier models have been previously considered by Fan et al. (1996), Martins-Filho and Yao (2007), Kumbhakar et al. (2007), among others. However, to the best of our knowledge, all of the above papers assumed that the regressors are exogenous. When some or all of the regressors are endogenous, their proposed estimation approaches are no longer valid and need to be modified. An issue that arises from nonparametric estimation and that is specific to stochastic frontier models is that economic theory leads to shape restrictions such as monotonicity, concavity, symmetry and linear homogeneity of the estimated function. Thus, it is important to incorporate these restrictions in nonparametric estimation.

In the context of parametric stochastic frontier models, the problem of endogeneity has been addressed by Kutlu (2010), Tran and Tsionas (2013, 2015), Griffiths and Hajargasht (2016), Karakaplan and Kutlu (2015), Amsler et al. (2016, 2017); Kutlu and Tran (2019), providing more references. In the context of nonparametric regression estimation with shape restrictions and exogenous regressors, some recent works include Hall and Huang (2001), Henderson and Parmeter (2009), Henderson et al. (2012), Du et al. (2013), Sun (2015) and Malikov et al. (2016a), just to name a few. The only paper that we are aware of that allows for endogenous regressors in a nonparametric setting is Freyberger and Horowitz (2015) where they propose monotone restrictions to identify an unknown regression with endogenous discrete covariates under the assumption that the discrete instruments have fewer mass points than covariates.

In this paper, we address the problem of endogeneity of regressors and shape restrictions in estimation of nonparametric stochastic frontier models. We propose several estimation procedures. Our first approach is to extend the nonparametric simultaneous equation models of Su and Ullah (2008) to a stochastic frontier framework where we impose shape restrictions using the method of Du et al. (2013). We propose a three-step estimation algorithm which provides a consistent estimator of the frontier and error variances and can be used to obtain inefficiency predictions. Our second approach is to use a local specification of a parametric frontier in the spirit of Gozalo and Linton (2000) so that the shape restrictions are satisfied locally, and we construct a similar

---

<sup>1</sup> The term “nonparametric” used in this paper refers specifically to the assumptions on the functional form of the frontier. That is, no specific functional form of the frontier is assumed.

three-step estimation procedure as in the first approach. Our Monte Carlo results show that the two approaches perform equally well in finite samples.

Additionally, we extend the smooth-coefficient stochastic frontier models of Sun and Kumbhakar (2013) to allow for endogenous regressors and develop a two-step estimation procedure for these models. Our approach here is different from the recent work by Kumbhakar et al. (2016) where they consider the estimation of a smooth-coefficient stochastic frontier production system.

The rest of the paper is organized as follows: Section 2 presents the nonparametric stochastic frontier model with endogenous regressors and provides detailed discussion of the two estimation approaches. A smooth-coefficient stochastic frontier model with endogenous regressors and the two-step estimation algorithm are presented in Sect. 3. Section 4 presents the Monte Carlo simulations to assess the performance of the proposed estimators discussed in Sects. 2 and 3. An empirical application from the US banking sector is given in Sect. 5. Section 6 concludes.

## 2 Nonparametric stochastic frontier model

First, we consider a nonparametric stochastic frontier model (SFM) with endogenous regressors in the simultaneous equation framework of the form:

$$\begin{cases} Y_i = g(X_i, Z_{1i}) + v_i - u_i, \\ X_i = h(Z_i) + \eta_i, \quad Z_i = (Z'_{1i}, Z'_{2i})', \end{cases} \tag{1}$$

where  $Y_i$  is an observable scalar random variable representing output of firm  $i$ ,  $g(\cdot)$  denotes the true, unknown structural frontier,  $X_i$  is  $d \times 1$  vector of inputs that are endogenous,  $Z_{1i}$  and  $Z_{2i}$  are  $l_1 \times 1$  and  $l_2 \times 1$  (where  $l_2 \geq d$ ) vectors of exogenous inputs and instrumental variables, respectively,  $h(\cdot) = (h_1(\cdot), \dots, h_d(\cdot))'$  is a  $d \times 1$  vector of functions of the instruments  $Z_i$ ,  $v_i$  is a two-sided symmetric random noise, and  $u_i$  is a one-sided error representing technical inefficiency of firm  $i$ . We are interested in estimating  $g(\cdot)$  consistently so that it satisfies the requisite axioms of production (or cost), and in using this estimate to predict the firm-specific inefficiency levels.

In this paper, we assume that  $u_i$  is independent of  $\xi_i = (v_i, \eta_i)'$  and that  $E(\varepsilon_i|Z_i, \eta_i) = E(\varepsilon_i|\eta_i)$ , where  $\varepsilon_i = v_i - u_i$ , so that  $\varepsilon_i$  is uncorrelated with  $Z_i$  given  $\eta_i$ ; we further assume that  $v_i \sim N(0, \sigma_v^2)$ ,  $u_i \sim N^+(0, \sigma_u^2)$  and, conditional on  $Z_i, \xi_i \sim N(0, \Omega)$ , where  $\Omega = \begin{pmatrix} \sigma_v^2 & \Omega_{v\eta} \\ \Omega_{\eta v} & \Omega_{\eta\eta} \end{pmatrix}$ . The assumption that the inefficiency term  $u_i$  is independent of the other error terms can be relaxed by introducing a copula function to model the joint distribution of  $(u_i, \xi_i)$  similarly to Amsler et al. (2016). Other assumptions on the marginal distributions of  $u_i$  and  $\xi_i$ , and/or allow for  $u_i$  to depend on a set of environmental variables can also be considered.<sup>2</sup> However, these extensions are beyond the scope of this paper and we leave them for future exploration.

<sup>2</sup> See, for example, Kumbhakar et al. (1991), Huang and Liu (1994), Battese and Coelli (1995), Caudill et al. (1995), Wang (2002) and Amsler et al. (2014) and the references therein.

Under these assumptions, the density of the endogenous variables conditional on the instruments can be decomposed as follows:

$$f(Y_i, X_i|Z_i) = f(Y_i|X_i, Z_i)f(X_i|Z_i). \tag{2}$$

Following the derivations of Amsler et al. (2016), we have:

$$f_{\varepsilon, \eta}(\varepsilon_i, \eta_i) = \text{const} \times |\Omega_{\eta\eta}|^{-1} \exp\left(-\frac{1}{2}\eta_i' \Omega_{\eta\eta}^{-1} \eta_i\right) \sigma^{-1} \phi\left(\frac{\varepsilon_i - \mu_{ci}}{\sigma}\right) \Phi\left(\frac{-\lambda(\varepsilon_i - \mu_{ci})}{\sigma}\right) \tag{3}$$

where  $\eta_i = X_i - h(Z_i)$ ,  $\varepsilon_i = Y_i - g(X_i, Z_{1i})$ ,  $\mu_{ci} = \Omega_{v\eta} \Omega_{\eta\eta}^{-1} \eta_i$ ,  $\lambda = \frac{\sigma_u}{\sigma_c}$ ,  $\sigma_c^2 = \sigma_v^2 - \Omega_{v\eta} \Omega_{\eta\eta}^{-1} \Omega_{v\eta}$ ,  $\sigma^2 = \sigma_u^2 + \sigma_c^2$ ,  $\phi(\cdot)$  and  $\Phi(\cdot)$  are, respectively, the standard normal density and CDF. Then, the log-likelihood function is given by:

$$\ln L = \ln L_1 + \ln L_2 \tag{4}$$

where

$$\begin{aligned} \ln L_1 = & -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - g(X_i, Z_{1i}) - \Omega_{v\eta} \Omega_{\eta\eta}^{-1} \eta_i)^2 \\ & + \sum_{i=1}^n \ln \Phi\left(\frac{-\lambda(Y_i - g(X_i, Z_{1i}) - \Omega_{v\eta} \Omega_{\eta\eta}^{-1} \eta_i)}{\sigma}\right) \end{aligned} \tag{5}$$

and

$$\ln L_2 = -\frac{n}{2} \ln |\Omega_{\eta\eta}| - \frac{1}{2} \sum_{i=1}^n (X_i - h(Z_i))' \Omega_{\eta\eta}^{-1} (X_i - h(Z_i)). \tag{6}$$

Unlike Amsler et al. (2016), this likelihood contains nonparametric terms.

### 2.1 Estimation: semiparametric LIML

Since the functions  $g(\cdot)$  and  $h(\cdot)$  are unknown, the log-likelihood function given in (4)–(6) is not feasible to maximize in practice. First, note that since  $u$  is independent of  $\eta$  and uncorrelated with  $Z$ , we have from (1):

$$\begin{aligned} E(Y|X, Z, \eta) &= g(X, Z_1) + E[(v - u)|X, Z, \eta] \\ &= g(X, Z_1) + E[(v - u)|h(Z) + \eta, Z, \eta] \\ &= g(X, Z_1) + E[(v - u)|Z, \eta] \\ &= g(X, Z_1) + E(v|Z, \eta) - E(u|Z, \eta) \\ &= g(X, Z_1) + E(v|\eta) - E(u) \end{aligned}$$

$$\begin{aligned}
 &= g^*(X, Z_1) + E(v|\eta) \\
 &= g^*(X, Z_1) + \Omega_{v\eta}\Omega_{\eta\eta}^{-1}\eta,
 \end{aligned}
 \tag{7}$$

where the second term on the right-hand side of the last equality in (7) comes from our assumption of joint normality of  $v$  and  $\eta$ , and where  $g^*(X, Z_1) = g(X, Z_1) - \mu_u$ , with  $\mu_u = E(u) = \sqrt{\frac{2}{\pi}}\sigma_u$ . Thus, from the law of iterated expectations, we have

$$E(Y|X, Z_1, \eta) = m(X, Z_1, \eta) = g^*(X, Z_1) + \Omega_{v\eta}\Omega_{\eta\eta}^{-1}\eta,
 \tag{8}$$

and Eq. (8) resembles a standard partially linear regression model. In the spirit of Fan et al. (1996), if we can obtain consistent estimates of  $g^*(X, Z_1)$ ,  $\Omega_{v\eta}$  and  $\Omega_{\eta\eta}$  then by replacing  $g(X, Z_1)$  by  $g^*(X, Z_1) + \mu_u$  and using the estimates of  $\Omega_{v\eta}$ ,  $\Omega_{\eta\eta}$  in (5), the remaining parameters can be obtained by maximizing (4). However, the frontier function  $g(X, Z_1)$  must satisfy relevant shape restrictions such as monotonicity and concavity for it to be considered a production function. Consequently, these restrictions must be imposed in the estimation of  $g(X, Z_1)$ .

In the context of nonparametric simultaneous equation models, Su and Ullah (2008) discuss how to obtain a consistent estimator of the unrestricted function  $g^*(X, Z_1)$  based on a three-step estimation procedure. Thus, one approach is to extend their estimation procedure to partially linear model that allows for resulting frontier to satisfy these restrictions. We call this approach *constrained semiparametric limited information maximum likelihood* (CSPLIML) method. Alternatively, we can follow Gozalo and Linton (2000) (see also McManus 1994) and approximate the function  $g(X, Z_1)$  locally by a parametric production function that satisfied these restrictions. That is,  $g(X, Z_1) \simeq m(X, Z_1, \alpha)$ . A natural candidate for  $m(\cdot, \cdot, \alpha)$  would be the Cobb–Douglas production function, albeit other production functions such as constant elasticity of substitution (CES) or Leontief can also be used.<sup>3</sup> We call this approach *local LIML* (LLIML). We now provide a detailed discussion of the two approaches.

### 2.1.1 CSPLIML

To obtain the constrained estimator of  $g^*(X, Z_1)$ , let  $W = (X, Z_1)$  and  $Q$  be a deterministic weighting function such that  $\int_{\mathfrak{R}^d} dQ(\xi) = 1$ , and  $q$  be the density of  $Q$  with respect the Lebesgue measure in  $\mathfrak{R}^d$ . Then, the constrained three-step procedure can be constructed as follows.

*Step 1* Obtain a consistent estimator of  $h(Z_i)$  by local linear (or general local polynomial) smoothing of  $X_i = (X_{1i}, \dots, X_{di})'$  on  $Z_i$  with kernel  $K_1$  and bandwidth sequence  $b_1(n)$ . Denote the estimates of  $h(Z_i)$  by  $\hat{h}(Z_i) = (\hat{h}_1(Z_i), \dots, \hat{h}_d(Z_i))'$ , and calculate the estimated residuals  $\hat{\eta}_i = (\hat{\eta}_{i1}, \dots, \hat{\eta}_{id})'$  where  $\hat{\eta}_{ij} = X_{ij} - \hat{h}_j(Z_i)$  for  $j = 1, \dots, d$ . In addition, compute the estimate of  $\Omega_{\eta\eta}$  by  $\hat{\Omega}_{\eta\eta} = n^{-1} \sum_{i=1}^n \hat{\eta}_i \hat{\eta}_i'$ ,

<sup>3</sup> In the parametric stochastic frontier literature, the Cobb–Douglas specification has been used most frequently in practice.

and let  $\hat{\eta}_i^* = \hat{\Omega}_{\eta\eta}^{-1} \hat{\eta}_i$ . Note that this step is equivalent to maximizing the log-likelihood function (6).

*Step 2* Obtain a consistent restricted estimator of  $m(W, \hat{\eta}) = g^*(W) + \Omega_{v\eta} \hat{\eta}^*$  by constrained profile likelihood method (or more precisely, constrained profile least squares in the current context) with the corresponding kernel  $K_2$  and bandwidth sequence  $b_2(n)$ . In order to describe the estimator in detail, we need more notation. Let  $A$  be an  $(n \times n)$  matrix with elements  $A_{ij} = A_i(W_j)$ , where  $A_i(w)$  is the kernel weight (e.g., for a local constant estimator,  $A_i(w) = K_{b_2}(W_i - w) / \sum_{i=1}^n K_{b_2}(W_i - w)$ ), and let  $\mathbf{p} = (p_1, \dots, p_n)$  and  $\mathbf{p}_u = (\frac{1}{n}, \dots, \frac{1}{n})$ . In the spirit of Du et al. (2013) and Parmeter and Racine (2013), the constrained kernel smoothing estimator is based on the solution  $(\hat{p}_1, \dots, \hat{p}_n)$  of a standard quadratic programming problem in which the objective function  $D(\mathbf{p}) = (\mathbf{p} - \mathbf{p}_u)'(\mathbf{p} - \mathbf{p}_u)$  is minimized subject to the relevant monotonicity and concavity constraints, as initially proposed by Hall and Huang (2001) in a univariate, single-constraint setting. Specifically, let  $Y = (Y_1, \dots, Y_n)$ ,  $W = (W_1', \dots, W_n')'$ ,  $W^* = (\mathbf{I}_n - A)W$  and  $Y^* = (\mathbf{I}_n - A)Y$ . Further denote  $\hat{\Omega}_{v\eta} = (W^{*'}W^*)^{-1}W^{*'}Y^*$  and  $\tilde{Y}_i = Y_i - \hat{\Omega}_{v\eta} \hat{\eta}_i^*$ . Then, the monotonicity and concavity constraints we impose are of the form:

- (i)  $\sum_{i=1}^n p_i A_i(w) \tilde{Y}_i - \tilde{Y}_i \geq 0$
- (ii)  $\sum_{i=1}^n p_i A_i^{(s_1)}(w) \tilde{Y}_i \geq 0$ , for  $s_1 \in S_1$ ,
- (iii)  $\sum_{i=1}^n p_i A_i^{(s_2)}(w) \tilde{Y}_i \leq 0$  for  $s_2 \in S_2$ ,

where  $S_1 = \{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, \dots, 1)\}$ ;  $S_2 = \{(2, 0, \dots, 0), (0, 2, \dots, 0), \dots, (0, \dots, 2)\}$ , and  $A_i^{(sr)}(w)$  is the  $r$ th order partial derivative of  $A_i(w)$  with respect to the  $j$ -th element of  $w$ ,  $j = 1, \dots, d + l_1$ . Given the solution  $(\hat{p}_1, \dots, \hat{p}_n)$ , we can obtain the restricted estimator of  $g^*(w)$  as  $\hat{g}^*(w) = \sum_{i=1}^n \hat{p}_i A_i(w) \tilde{Y}_i$ .

*Step 3* Once the estimates  $\hat{g}^*(X, Z_1)$  are obtained, by replacing  $g(X, Z_1)$  in (5) with  $\hat{g}^*(X, Z_1) + \mu_u$ , the remaining variance parameters  $(\sigma_v^2, \sigma_u^2)$  can be obtained by maximizing Eq. (5). Let  $(\hat{\sigma}_v^2, \hat{\sigma}_u^2)$  denote the resulting estimates, the structural frontier  $g(X, Z_1)$  can be estimated by  $\hat{g}(X, Z_1) = \hat{g}^*(X, Z_1) + \hat{\mu}_u$  where  $\hat{\mu}_u = \hat{\sigma}_u \sqrt{2/\pi}$ .

The following remarks are worth noting:

**Remark 1** As is the case with all nonparametric estimations, implementing the estimator  $\hat{g}^*(\cdot)$  (and hence  $\hat{g}(\cdot)$ ) requires specification of a kernel function and a method of choosing the values for the two bandwidth parameters  $b_1(n)$  and  $b_2(n)$ . Let  $p_1$  and  $p_2$  denote the order of the local polynomials in Steps 1 and 2, respectively. As Su and Ullah (2008) pointed out for the case  $p_2 = 1$ , as long as Assumption A5 in their paper is satisfied, the bandwidth parameter  $b_1$  does not affect the asymptotic distribution of  $\hat{g}(\cdot)$ , while  $b_2$  does. They suggest a systematic method for choosing  $b_2$  and a rule of thumb for choosing  $b_1$ . We follow these suggestions. For the kernel function, we use a product of second-order univariate kernels, and the rule of thumb for selecting  $b_1$  is given by  $\hat{b}_1 = \hat{b}_2 n^{-(\alpha_2 - \alpha_1)/(2\zeta)}$  where  $\alpha_1 = \max \left\{ \frac{p_2 + 1}{p_1 + 1}, \frac{p_2 + 3}{2(p_1 + 1)} \right\}$ ,  $\alpha_2 = \frac{p_2 + d + l_1 - 1}{l_1 + l_2}$ ,  $\zeta = 2(p_2 + 1) + d + l_1$ , and  $\hat{b}_2$  is the optimal “plug-in” bandwidth which is given by  $\hat{b}_2 = \left\{ \frac{(\gamma_{02})^{d+l_1} (d+l_1) \tilde{c}_2}{(\gamma_{21})^2 \tilde{c}_1} \right\}^{1/(4+d+l_1)} \times n^{-1/(4+d+l_1)}$ , where  $\gamma_{ij} = \int_{\mathfrak{R}} t^i K_2(t)^j dt$ , for

$i = 0, 2$  and  $j = 1, 2$ ;  $\tilde{c}_1$  and  $\tilde{c}_2$  are defined in equations (2.11) and (2.13) of Su and Ullah (2008).

**Remark 2** <sup>4</sup> The estimation of  $E(Y|X, Z_1, \eta)$  requires an estimate of  $h(Z_i)$ . However, it is possible in principle to use the information contained in the constraints on  $E(Y|X, Z_1, \eta)$  during the Step 1 estimation of  $h(Z_i)$ . This may provide an improvement to the Step 1 estimator, but its asymptotics does not depend on these constraints and so any potential gains would be limited to finite samples. Also, for large  $n$ , the computation in Step 2 can be cumbersome and expensive due to the localization of the entire sample. To reduce this computation burden, Henderson and Parmeter (2015) show that by using a smaller (random) sample of, say, 20 to 30 percent of  $n$ , there is almost no loss in the estimation precision. We follow the suggestion of Henderson and Parmeter (2015) in our Monte Carlo simulations and empirical application below.

### 2.1.2 LLIML

In this section, we describe an alternative estimation procedure for  $m(W, \hat{\eta})$  in Step 2 of CSPLIML described above. In what follows, Step 1 and Step 3 are the same as in CSPLIML but Step 2 is different.

Now, instead of directly imposing the constraints in the estimation of  $g^*(W)$  in Step 2, we follow Gozalo and Linton (2000) and use an anchoring parametric specification in which all the theoretical restrictions are satisfied in a local approximation of  $g^*(W)$ . In this paper, we use the Cobb–Douglas production function as the anchoring model. McManus (1994) showed that the nonparametric estimator based on the local version of the Cobb–Douglas specification is efficient among the class of all linear nonparametric estimators and nearly efficient among all nonparametric estimators. McManus (1994) provides a detailed discussion of this point and an empirical application.

We use the following approximation:  $g^*(W, \alpha) \simeq \alpha_0 W_1^{\alpha_1} \dots W_p^{\alpha_p}$  or its log-linear version,  $g^*(W, \alpha) \simeq \ln \alpha_0 + \sum_{j=1}^p \alpha_j \ln W_j$ , where  $p = (d + l_1)$ . Note that the Cobb–Douglas functional form explicitly imposes the restriction that  $g^*(0, \alpha) = 0$  and the monotonicity and concavity are automatically satisfied. Moreover, one can also impose a constant return to scale, if desired, by writing  $\alpha_p = 1 - \sum_{j=2}^{p-1} \alpha_j$ .

The estimate of  $g^*(W, \alpha)$  can be obtained by minimizing the following criterion:

$$S_n(w, \alpha) = n^{-1} \sum_{i=1}^n \{Y_i - g^*(W_i, \alpha) - \Omega_{v\eta} \hat{\eta}_i^*\}^2 K_H(W_i - w) \tag{9}$$

where  $K_H(\zeta) = \det(H)^{-1} K(H^{-1}\zeta)$ , with  $H$  being a  $(d \times d)$  nonsingular bandwidth matrix, while  $K(\cdot)$  is a real value kernel function. The estimate of  $g^*(W)$  is  $\hat{g}^*(w) = g^*(w, \hat{\alpha}(w))$ . Note that for this estimator, the optimal bandwidth can be selected by cross-validation and the bandwidth for the first stage can be selected using the rule of thumb as discussed in Sect. 2.1.1.

<sup>4</sup> We thank the Special Issue editors for bringing these points to our attention.

## 2.2 Predicting inefficiency

Once the estimates of the frontier and the variances parameters are obtained, a popular predictor of  $u_i$  is  $\hat{u}_i = E(u_i|\epsilon_i)$  as suggested by Jondrow et al. (1982). Amsler et al. (2016) improve upon the predictor by pointing out that given the reduced form equation, a better predictor of  $u_i$  can be constructed based on the additional information contained in  $\eta_i$ , namely  $\hat{u}_i = E(u_i|\epsilon_i, \eta_i)$ . That is, albeit  $\eta_i$  is independent of  $u_i$ , it is correlated with  $v_i$  and hence informative about  $v_i$ . Following Amsler et al. (2016), we estimate  $\hat{u}_i$  as follows:

$$\hat{u}_i = E(u_i|\epsilon_i, \eta_i) = E(u_i|\tilde{\epsilon}_i) = \hat{\sigma}_* \left[ \frac{\phi(\hat{\lambda}\hat{\tilde{\epsilon}}_i/\hat{\sigma})}{1 - \Phi(\hat{\lambda}\hat{\tilde{\epsilon}}_i/\hat{\sigma})} - \frac{\hat{\lambda}\hat{\tilde{\epsilon}}_i}{\hat{\sigma}} \right] \quad (10)$$

where  $\hat{\tilde{\epsilon}}_i = \hat{\epsilon}_i - \hat{\mu}_{ci}$  and  $\hat{\sigma}_*^2 = \hat{\sigma}_u^2 \hat{\sigma}_c^2 / \hat{\sigma}^2$ . Equation (10) provides a better estimate than the original formulation,  $E(\hat{u}_i|\epsilon_i)$ , due to the fact that  $\sigma_c^2 < \sigma_v^2$ . However, the main trade-off or disadvantage is that normality of errors in the reduced form equations is assumed to be correctly specified.

Simar et al. (2017) propose an ingenious alternative by estimating  $E(u_i|Z_i, X_i)$  using  $\hat{\sigma}_u^2$ , a nonparametric estimator of  $\sigma_u^2$ , which in their case is allowed to depend on  $Z_i$  and  $X_i$ .<sup>5</sup> An attractive feature of this estimator is that it conditions on observed values of covariates, rather than on the unobserved errors. They obtain  $\hat{\sigma}_u^2(Z_i, X_i)$  from local estimators of second and third-order (conditional) moments of  $v - u + E(u|X, Z)$ , which is a zero-mean error. The higher-order moments of this error can be expressed in terms of (conditional) moments of  $u$  which are well known functions of  $\sigma_u^2(Z_i, X_i)$  if  $u$  is half-normal.

This estimator does not require normality of  $v$ . Under the additional assumption that  $v$  is normal, Simar et al. (2017) mention the option of using a heteroskedastic version of Jondrow et al. (1982). In this case, the predictor of  $u_i$  becomes  $E(u_i|\epsilon_i, Z_i X_i)$  and we compute it using the formula for  $E(u_i|\epsilon_i)$  from Jondrow et al. (1982) after replacing  $\sigma_u^2$  with the local estimator  $\hat{\sigma}_u^2(Z_i, X_i)$ .

In a similar manner, we can further improve upon the estimators proposed by Simar et al. (2017) by considering a heteroskedastic version of Amsler et al. (2016). In this case, the predictor of  $u_i$  can be written as  $E(u_i|\epsilon_i, \eta_i, Z_i X_i)$ , where we use  $\hat{\sigma}_u^2(Z_i, X_i)$  in place of  $\hat{\sigma}_u^2$  in (11).

## 3 Semiparametric SFM: smooth coefficients frontier

In Sect. 2, we considered nonparametric specifications of the frontier which are quite flexible and robust to misspecification of the functional form. However, the main drawback is that when the number of regressors is large, the curse of dimensionality prohibits precise estimation in such models. In this section, we propose an alternative class of semiparametric frontiers that are still flexible and yet, more resistant to the curse of dimensionality.

<sup>5</sup> We thank an anonymous referee for pointing out this estimator to us.



Consider the following semiparametric smooth coefficients (SC) stochastic frontier model:

$$Y_i = \alpha(Z_i) + X_i' \beta(Z_i) + v_i - u_i, \quad i = 1, \dots, n, \tag{11}$$

where  $Y_i$ ,  $X_i$ ,  $v_i$  and  $u_i$  are defined as in previous section;  $Z_i$  is a  $p \times 1$  vector of environmental factors,  $\alpha(Z_i)$  is an intercept and  $\beta(Z_i)$  is a  $d \times 1$  vector; both are assumed to be unknown but smooth functions of  $Z_i$ . As before, we allow for some or all components in  $X_i$  to be correlated with  $v_i$  but assume that  $Z_i$  are uncorrelated with  $v_i$  and  $u_i$ , and  $v_i$  and  $u_i$  are independent. Moreover, we assume that  $v_i \sim$  i.i.d.  $N(0, \sigma_v^2)$ ,  $u_i \sim$  i.i.d.  $N^+(0, \sigma_u^2(Z_i))$ , where  $\sigma_u(Z_i) = \exp(\delta_0 + Z_i' \delta_1)$ . Model (11) is essentially an extension of the model considered by Sun and Kumbhakar (2013) by allowing for the endogeneity of  $X_i$ .

It is important to note that in practice, the assignment of relevant variables  $X_i$  and  $Z_i$  is the practitioner's prerogative, and it may be done on the basis of practical convenience or on the basis of economic theory. Nevertheless, our results do not impose any restrictions on the relationship between  $X_i$  and  $Z_i$ . That is,  $X_i$  may or may not be different from or unrelated to  $Z_i$ .

To estimate the unknown coefficient functions and the variance parameters, we suggest in the spirit of Sun and Kumbhakar (2013), a two-step estimation approach. We now provide details of this estimator.

Let  $\epsilon_i = v_i - (u_i - E(u_i|Z_i)) = v_i - u_i^*$ , where the definition of  $u_i^*$  is apparent, and under our assumption,  $E(u_i|Z_i) = \sqrt{2/\pi} \sigma_u(Z_i)$ . Then, (11) can be expressed as:

$$Y_i = \alpha^*(Z_i) + X_i' \beta(Z_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{12}$$

where  $\alpha^*(Z_i) = \alpha(Z_i) - \sqrt{2/\pi} \sigma_u(Z_i)$  and  $E(\epsilon_i) = 0$ . Define  $\theta(Z_i) = [\alpha^*(Z_i), \beta(Z_i)']'$  and  $\tilde{X}_i = [1, X_i']'$ , then (12) can be compactly written as:

$$Y_i = \tilde{X}_i' \theta(Z_i) + \epsilon_i, \quad i = 1, \dots, n. \tag{13}$$

Thus, Eq. (13) resembles the smooth-coefficient stochastic frontier model of Sun and Kumbhakar (2013). However, due to the endogeneity of  $X_i$ , their approach cannot be directly used. To obtain consistent estimators of the unknown functions and parameters of the model, we suggest a two-step estimation procedure where in the first two steps, a local generalized method of moments (LGMM) is used to obtain consistent estimator for the unknown coefficient function  $\theta(Z_i)$ ; while in the second step, the MLE is used to obtain the remaining unknown parameters. A detailed description of the two-step estimation procedure is given below.

*Step 1: LGMM Estimation of  $\theta(Z_i)$ .*

Suppose there exists an  $1 \times l$  (where  $l \geq d$ ) vector of instruments  $W_i$  (including a constant) such that  $W_i$  are uncorrelated with  $\epsilon_i$  so that  $E(W_i' \epsilon_i) = 0$ . Assume that the elements of  $\theta(Z_i)$  are twice continuously differentiable in the neighborhood of  $z$ .

Then, for  $Z_i$  in the neighborhood of  $z$ , we have the following locally kernel-weighted orthogonality conditions:

$$E \left[ W_i' K_{hi}(z) (Y_i - \tilde{X}'_i \theta(Z_i)) \right] = 0, \quad i = 1, \dots, n, \tag{14}$$

where  $K_{hi}(z) = \prod_{j=1}^p h_j^{-1} k((Z_i - z)/h_j)$  is a product kernel function in which  $k(\cdot) \geq 0$  is a bounded univariate symmetric kernel satisfying the conditions  $\int k(\psi) d\psi = 0$ ,  $\int \psi^2 k(\psi) d\psi = c_1 > 0$  and  $\int k^2(\psi) d\psi = c_2 > 0$ , and  $h = (h_1, \dots, h_p)'$  is a  $(p \times 1)$  vector of bandwidths. Thus, Eq. (14) provides the moment restrictions to construct the following local GMM criterion function:

$$J_n(\theta(z)) = \sum_{i=1}^n \left[ \{Y_i - \tilde{X}'_i \theta(Z_i)\}' K_{hi}(z) W_i \right] \left[ W_i' K_{hi}(z) \{Y_i - \tilde{X}'_i \theta(Z_i)\} \right]. \tag{15}$$

Minimizing (15) with respect to  $\theta(\cdot)$  yields the following local GMM estimator:

$$\hat{\theta}(z) = \left( \sum_{i=1}^n \tilde{X}'_i \Psi_{hi}(z) \tilde{X}_i \right)^{-1} \left( \sum_{i=1}^n \tilde{X}'_i \Psi_{hi}(z) Y_i \right) \tag{16}$$

where  $\Psi_h(z) = K_{hi}(z) W_i W_i' K_{hi}(z)$ . Let  $\hat{\epsilon}_i = Y_i - \tilde{X}'_i \hat{\theta}(Z_i)$  be the estimated residuals from (13), then the remaining parameters, namely  $\gamma = (\sigma_v, \delta')'$  where  $\delta = (\delta_0, \delta_1)'$ , can be estimated using the ML procedure below.

*Step 2: ML Estimation of  $\gamma$ .* First, recall that  $\epsilon_i$  was defined as  $\epsilon_i = v_i - u_i^* = v_i - u_i + \sqrt{2/\pi} \sigma_u(Z_i)$ , where  $\sigma_u(Z_i) = \exp(\delta_0 + Z'_i \delta_1)$ , and hence can be re-expressed as:

$$\epsilon_i = \sqrt{2/\pi} \exp(\delta_0 + Z'_i \delta_1) + v_i - \exp(\delta_0 + Z'_i \delta_1) \tilde{u}_i, \tag{17}$$

where  $\tilde{u}_i \sim N^+(0, 1)$ . Let  $\sigma_i = \sigma_v + \sigma_u(Z_i) = \sigma_v + \exp(\delta_0 + Z'_i \delta_1)$ , and  $\lambda_i = \sigma_u(Z_i)/\sigma_v$ . Then, under our assumptions, the log-likelihood function associated with (17) is given by:

$$\ln L(\gamma) = \text{Constant} - \frac{1}{2} \sum_{i=1}^n \ln(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^n \frac{\hat{\epsilon}_i^2}{\sigma_i} + \sum_{i=1}^n \ln \Phi \left( -\frac{\lambda_i \hat{\epsilon}_i}{\sigma_i} \right). \tag{18}$$

The MLE of  $\gamma$  is then given by:

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \ln L(\gamma), \tag{19}$$

where  $\Gamma$  is the parameter space of  $\gamma$  which is assumed to be compact. The estimators of  $\sigma_u(Z_i)$ ,  $\lambda_i$  and  $E(u_i | Z_i)$  can be obtained as  $\hat{\sigma}_u(Z_i) = \exp(\hat{\gamma}_0 + Z'_i \hat{\gamma}_1)$ ,  $\hat{\lambda}_i = \hat{\sigma}_u(Z_i)/\hat{\sigma}_v$

and  $\hat{E}(u_i|Z_i) = \sqrt{2/\pi}\hat{\sigma}_u(Z_i)$ , respectively. The estimator of  $\hat{E}(u_i|Z_i)$  is needed to identify the intercept term in (12).

We make the following remarks:

**Remark 3** The model given in (13) is identical to the model considered by Tran and Tsonas (2009) and hence, the asymptotic properties of the LGMM estimator  $\hat{\theta}(Z_i)$  can be drawn directly from the results of Theorem 1 in their paper. As for the MLE estimator  $\hat{\gamma}$ , it can also be shown to be consistent and asymptotically normal.

**Remark 4** It is clear from (13) that the variance structure of  $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$  is heteroskedastic since  $\text{Var}(\epsilon) = \Omega = \sigma_v^2 I_n + \Sigma_u$  where  $\Sigma_u = \text{diag}(\sigma_u^2(Z_1), \dots, \sigma_u^2(Z_n))$ . Thus, it is tempting to construct an improved (i.e., more efficient) LGMM estimator of  $\theta(Z_i)$  by modifying the objective function in (15) to account for this variance structure. However, such a LGMM estimator can be problematic for the following two reasons.<sup>6</sup> First, as the literature has shown (see, e.g., Martins-Filho and Yao 2009; Su et al. 2013; Parmeter and Racine 2019), incorporating the variance-covariance structure in a nonparametric regression model is a delicate balance of bias and efficiency that does not always lead to improvements even in large samples. Second, as argued by Lin and Carroll (2000) for a typical random effects panel data models, when a standard kernel-based estimator is used in estimating the regression, it is often better to use the “working independence” approach by ignoring the correlation structure within a cluster. Thus, additional work on this topic will be fruitful and we leave it for future research.

We now turn to estimation of the marginal effects of inefficiency determinants and inefficiency predictions.

Recall that  $E(u_i|Z_i) = \mu_u(Z_i) = \sqrt{2/\pi} \exp(\delta_0 + Z_i' \delta_1)$ ; hence, the marginal effect of  $\mu_u(Z_i)$  with respect to the  $j^{\text{th}}$  element of  $Z_i$  (which we will denote as  $z_{ij}$ ) can be estimated as:

$$\hat{\delta}_j = \frac{\partial \mu_i(Z_i)}{\partial z_{ij}} = \sqrt{2/\pi} \hat{\delta}_{1j}, \tag{20}$$

where  $\hat{\delta}_{1j}$  is the MLE estimate from Step 2.

Next, given the estimates of the model’s parameters, we compute the efficiency,  $\hat{\xi}_i$ , using the following Jondrow et al. (1982) approach to predict  $u_i$  conditional on  $\epsilon_i$ :

$$\hat{u}_i = \hat{E}(u_i|\epsilon_i) = \frac{\hat{\sigma}_v \hat{\sigma}_{ui}}{\hat{\sigma}_i} \left( \frac{\hat{\phi}_i}{1 - \hat{\Phi}_i} - \frac{\hat{\lambda}_i \hat{\epsilon}_i}{\hat{\sigma}_i} \right), \tag{21}$$

where  $\hat{\xi}_i = \exp(-\hat{u}_i)$ ,  $\hat{\sigma}_{ui} = \hat{\sigma}_u(Z_i)$ ,  $\hat{\phi}_i = \phi\left(-\frac{\hat{\lambda}_i \hat{\epsilon}_i}{\hat{\sigma}_i}\right)$  and  $\hat{\Phi}_i = \Phi\left(-\frac{\hat{\lambda}_i \hat{\epsilon}_i}{\hat{\sigma}_i}\right)$ .

As alternatives, it is possible to use the estimators of Simar et al. (2017) and the other improved estimators discussed in Sect. 2.2 to obtain  $\hat{\delta}_j$  and  $\hat{\xi}_i$  but we do not pursue these alternatives here.

<sup>6</sup> We would like to thank an anonymous referee for pointing these out.

## 4 Monte Carlo simulations

### 4.1 Data generating processes (DGPs)

To examine the finite sample performance of the proposed estimators, we conduct Monte Carlo experiments. We consider three data generating processes (DGPs):

*a. Nonparametric Frontier:*

$$\text{DGP 1: } \begin{cases} Y_i = g(X_i, Z_{1i}) = 1 + 0.5 \log(X_i + 1) + 0.5 \log(Z_{1i}) + v_i - u_i \\ X_i = Z_{2i} + \eta_i \end{cases}$$

$$\text{DGP 2: } \begin{cases} Y_i = g(X_i, Z_{1i}) = \log \left[ (0.5X_i^2 + 0.5Z_{1i}^2 + X_iZ_{1i})^{1/2} \right] + v_i - u_i \\ X_i = \log(0.2 + Z_{2i}^2) + \eta_i \end{cases}$$

where the errors  $v_i$  and  $\eta_i$  are generated as

$$\begin{pmatrix} v_i \\ \eta_i \end{pmatrix} \sim \text{i.i.d. } N \left( 0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Here,  $\rho = \{0.2, 0.5, 0.8\}$  indicate weak, medium and strong endogeneity, respectively. The one-sided error  $u_i$  is generated as  $u_i \sim \text{i.i.d. } N^+(0, \lambda^2 \sigma_c^2)$ , where  $\sigma_c^2 = 1 - \rho^2$  and we set  $\lambda = \{1.0, 2.0\}$ . The exogenous variable  $Z_{1i}$  and the instrument  $Z_{2i}$  are generated as i.i.d. each uniformly distributed on the interval  $[0, 2]$ . It can be easily verified that the above three DGPs satisfy the monotonicity and concavity conditions.

*b. Smooth Coefficient Frontier:*

$$\text{DGP 3: } Y_i = \alpha(Z_i) + \beta_1(Z_i)X_{i1} + \beta_2(Z_i)X_{i2} + v_i - u_i,$$

where  $\alpha(Z_i) = \exp(Z_i)$ ,  $\beta_j = \exp(a_j Z_i)$  for  $j = 1, 2$  with  $a_1 = 0.5$  and  $a_2 = 0.75$ . The term  $u_i$  is generated as  $u_i \sim N^+(0, \sigma_u^2(Z_i))$ , where  $\sigma_u^2(Z_i) = \exp(\delta Z_i)$ , and the values of  $\delta$  are implied by  $\lambda = \{1, 2\}$ . All the remaining variables are generated as in DGP 1. For each DGP, we consider two sample sizes,  $n = \{100, 200, 400, 800\}$ , and the number of replications we run is 500.

For comparison purposes, we also include, along with our proposed estimators, the estimators from the following scenarios for DGP 1 and DGP 2: (1) a correct specification of the frontier with endogeneity which we denote as MLE-CE, and this is our benchmark; (2) a correct specification of the frontier but ignoring endogeneity which we call MLE-C; and (3) the Fan et al. (1996) semiparametric estimator that ignores endogeneity which we label as FLW.

### 4.2 Simulation results

To assess the performance of the proposed estimators, we report, for DGP 1 and DGP 2, the following measures: (i) the average mean squared errors (AMSE) of the estimated CSPLIML and LLIML estimators over the number of replications; (ii) the ratio of

AMSE for each estimator, described at the end of previous subsection, relative to that of the proposed estimators over the realized values of  $X$  and  $Z_1$  and over the number of replications. The values of the ratios that are larger than 1 indicate superior AMSE performance of the proposed approaches; and (iii) the proportion of each constraint being violated. For DGP 3, we report only the average of MSE for the proposed two-step estimator over the number of replications.

For DGP 1 and 2, to obtain the estimator  $\hat{g}$ , we set  $p_1 = 3$  and  $p_2 = 1$ , and we use a product of the second-order Epanechnikov kernels standardized to have unit variance:

$$k_i(w) = \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}w^2\right) 1(|w| \leq \sqrt{5}), \quad i = 1, 2, 3.$$

The bandwidth sequences  $(b_1, b_2)$  are chosen as previously described. We also conduct experiments with different bandwidth choices with  $\hat{b}_1 = c\hat{b}_2n^{-(\alpha_2-\alpha_1)/(2\zeta)}$  where  $c$  is constant and  $c = \{0.5, 0.75, 1.25, 1.5, 2.0\}$ , and  $\hat{b}_2$  is defined as in Remark 1. Our results indicated that the estimates are similar for various values of  $c$ , with the value of  $c = 1$  providing the smallest AMSE. Consequently, we set  $c = 1$  throughout the simulations.<sup>7</sup> For DGP 3, the bandwidths are selected based on leave-one-out cross-validation.

The simulation results for DGPs 1 and 2 (nonparametric frontier) are reported in Tables 1 and 2. Note that the row ‘‘Violation’’ contains two entries where the top entry shows the proportion of points violating the monotonicity (m) and the bottom entry shows the proportion of points violating the concavity constraints (c) when we use the proposed CSPLIML and LLIML methods.<sup>8</sup>

First, we discuss the results from the nonparametric frontier. From Tables 1 and 2, we observe that the MSEs for the function  $g(\cdot)$  are very similar for both approaches, and the same is observed for the variance and correlation parameters. As the sample size doubles, the MSEs for  $g(\cdot)$  reduce and, for variance and correlation parameters, the MSEs decrease roughly to half indicating the root- $n$  consistency. The last row of Tables 1 and 2 shows the proportion of violations of the constraints. We observe that the CSPLIML method has more points violating both constraints than the LLIML method, albeit the proportions of both constraints violations are relatively small.

Next, to illustrate the usefulness of nonparametric specification of the frontier when some regressors are endogenous, Tables 3 and 4 provide the comparison of our proposed model/approach to some of the parametric models/approaches discussed at the end of Sect. 4.1 for DGP 1 and 2 (with  $\lambda = 2$  and  $\rho = 0.5$ ), respectively. Compared to the benchmark case, we see that our proposed estimators perform reasonably well in terms of MSEs, and as the sample size increases, our proposed semiparametric methods converge to the correctly specified parametric model with endogeneity for both DGPs considered (see MLE-CE column as  $n$  increases). In addition, as expected, our results also show that our proposed estimators outperform both the correctly spec-

<sup>7</sup> For conservation of space, we do not report these results here but they are available from the authors upon request.

<sup>8</sup> Let  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  be rank ordered values. We say that the function  $g(\cdot)$  violates the monotonically increasing condition at  $x(i)$  if  $g(x_{(i)}) < g(x_{(i-1)})$ . Moreover, we say that the function  $g(\cdot)$  violates the concavity condition at  $x_i$  if its Hessian matrix is positive definite.

**Table 1** MSE-DGP 1 (nonparametric frontier)

	$\lambda = 2$															
	$\lambda = 1$				$\rho = 0.2$				$\rho = 0.5$				$\rho = 0.8$			
	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML
<i>n</i> = 100																
<i>g</i>	0.357	0.362	0.339	0.352	0.225	0.264	0.165	0.175	0.144	0.170	0.110	0.112				
$\sigma_v$	0.258	0.271	0.277	0.283	0.155	0.192	0.140	0.153	0.097	0.101	0.075	0.078				
$\sigma_u$	0.268	0.245	0.257	0.240	0.168	0.180	0.151	0.153	0.080	0.009	0.062	0.074				
$\rho$	0.097	0.095	0.070	0.078	0.050	0.058	0.034	0.040	0.021	0.023	0.012	0.015				
Violation	(m) 0.047	0.000	0.045	0.000	0.043	0.000	0.044	0.000	0.042	0.000	0.045	0.000				
	(c) 0.035	0.000	0.037	0.000	0.035	0.000	0.035	0.000	0.030	0.000	0.032	0.000				
<i>n</i> = 200																
<i>g</i> ( $\cdot$ )	0.250	0.260	0.246	0.247	0.158	0.186	0.115	0.125	0.103	0.121	0.078	0.080				
$\sigma_v$	0.183	0.193	0.194	0.197	0.109	0.133	0.100	0.110	0.068	0.073	0.052	0.054				
$\sigma_u$	0.189	0.172	0.184	0.171	0.119	0.129	0.107	0.110	0.056	0.055	0.043	0.053				
$\rho$	0.069	0.069	0.049	0.056	0.035	0.041	0.024	0.028	0.015	0.016	0.008	0.010				
Violation	(m) 0.044	0.000	0.042	0.000	0.043	0.000	0.040	0.000	0.041	0.000	0.040	0.000				
	(c) 0.030	0.000	0.029	0.000	0.031	0.000	0.033	0.000	0.032	0.000	0.031	0.000				
<i>n</i> = 400																
<i>g</i> ( $\cdot$ )	0.187	0.188	0.165	0.167	0.121	0.128	0.087	0.095	0.071	0.081	0.059	0.061				
$\sigma_v$	0.140	0.138	0.132	0.134	0.082	0.091	0.076	0.079	0.048	0.053	0.038	0.040				
$\sigma_u$	0.136	0.132	0.127	0.128	0.087	0.094	0.074	0.080	0.040	0.041	0.033	0.036				
$\rho$	0.048	0.049	0.037	0.039	0.026	0.030	0.017	0.019	0.010	0.011	0.006	0.007				
Violation	(m) 0.030	0.000	0.035	0.000	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
	(c) 0.039	0.000	0.039	0.000	0.033	0.000	0.000	0.000	0.012	0.000	0.017	0.000				

Table 1 continued

	$\lambda = 2$															
	$\lambda = 1$				$\rho = 0.2$				$\rho = 0.5$				$\rho = 0.8$			
	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML		
$n = 800$																
$g(\cdot)$	0.177	0.171	0.100	0.101	0.080	0.087	0.050	0.058	0.043	0.043	0.040	0.037	0.036			
$\sigma_v$	0.065	0.062	0.055	0.057	0.044	0.048	0.029	0.032	0.018	0.018	0.016	0.011	0.012			
$\sigma_u$	0.069	0.062	0.060	0.063	0.052	0.057	0.022	0.027	0.020	0.020	0.024	0.013	0.014			
$\rho$	0.021	0.018	0.013	0.010	0.010	0.013	0.007	0.009	0.004	0.004	0.006	0.003	0.004			
Violation	(m) 0.040	0.000	0.030	0.000	0.030	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000			
	(c) 0.037	0.000	0.026	0.000	0.024	0.000	0.010	0.000	0.010	0.010	0.000	0.000	0.000			

**Table 2** MSE-DGP 2 (Nonparametric frontier)

	$\lambda = 2$															
	$\lambda = 1$				$\rho = 0.2$				$\rho = 0.5$				$\rho = 0.8$			
	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML
<i>n</i> = 100																
<i>g</i> (·)	0.351	0.341	0.331	0.326	0.300	0.315	0.315	0.212	0.214	0.205	0.194	0.174	0.165			
$\sigma_v$	0.277	0.264	0.222	0.235	0.225	0.237	0.237	0.085	0.080	0.066	0.067	0.051	0.062			
$\sigma_u$	0.267	0.275	0.240	0.248	0.226	0.217	0.217	0.077	0.077	0.063	0.065	0.050	0.051			
$\rho$	0.112	0.110	0.092	0.073	0.065	0.065	0.065	0.019	0.022	0.015	0.018	0.017	0.017			
Violation	(m) 0.043	0.000	0.045	0.000	0.046	0.000	0.000	0.047	0.000	0.045	0.000	0.043	0.000			
	(c) 0.032	0.000	0.032	0.000	0.032	0.000	0.000	0.033	0.000	0.036	0.000	0.032	0.000			
<i>n</i> = 200																
<i>g</i>	0.247	0.246	0.236	0.230	0.217	0.228	0.228	0.150	0.151	0.164	0.140	0.122	0.118			
$\sigma_v$	0.197	0.184	0.161	0.170	0.157	0.165	0.165	0.059	0.057	0.048	0.047	0.036	0.043			
$\sigma_u$	0.189	0.191	0.170	0.178	0.157	0.152	0.152	0.055	0.054	0.046	0.046	0.036	0.037			
$\rho$	0.081	0.076	0.065	0.051	0.047	0.046	0.046	0.014	0.016	0.011	0.012	0.012	0.012			
Violation	(m) 0.045	0.000	0.044	0.000	0.042	0.000	0.000	0.041	0.000	0.041	0.000	0.042	0.000			
	(c) 0.031	0.000	0.033	0.000	0.030	0.000	0.000	0.030	0.000	0.029	0.000	0.030	0.000			
<i>n</i> = 400																
<i>g</i> (·)	0.181	0.180	0.170	0.168	0.160	0.162	0.162	0.107	0.108	0.098	0.099	0.082	0.083			
$\sigma_v$	0.139	0.137	0.122	0.118	0.110	0.111	0.111	0.040	0.040	0.032	0.034	0.027	0.029			
$\sigma_u$	0.132	0.130	0.125	0.119	0.110	0.110	0.110	0.038	0.039	0.031	0.031	0.026	0.026			
$\rho$	0.057	0.056	0.049	0.036	0.032	0.034	0.034	0.010	0.011	0.008	0.009	0.008	0.009			
Violation	(m) 0.021	0.000	0.033	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
	(c) 0.030	0.000	0.030	0.028	0.010	0.015	0.015	0.018	0.000	0.010	0.000	0.000	0.000			



Table 2 continued

	$\lambda = 2$															
	$\lambda = 1$				$\rho = 0.2$				$\rho = 0.5$				$\rho = 0.8$			
	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML	CSPLIML	LLIML		
$n = 800$																
$g(\cdot)$	0.110	0.112	0.114	0.115	0.088	0.089	0.067	0.065	0.057	0.058	0.046	0.037	0.037	0.037		
$\sigma_v$	0.068	0.070	0.060	0.062	0.047	0.049	0.018	0.019	0.014	0.015	0.009	0.008	0.008	0.008		
$\sigma_u$	0.056	0.058	0.052	0.055	0.049	0.051	0.018	0.020	0.013	0.014	0.008	0.009	0.009	0.009		
$\rho$	0.022	0.024	0.020	0.022	0.019	0.020	0.010	0.012	0.008	0.008	0.007	0.008	0.008	0.008		
Violation	(m) 0.030	0.000	0.046	0.000	0.056	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	(c) 0.025	0.000	0.040	0.000	0.043	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000		

**Table 3** Ratio of AMSPE and MSE of Each Estimator Relative to CSPLIML and LLIML-DGP 1 ( $\lambda = 2$ ,  $\rho = 0.5$ )

	Relative to CSPLIML			Relative to LLIML		
	MLE-CE	MLE-C	FLW	MLE-CE	MLE-C	FLW
<i>n</i> = 100						
<i>g</i> (·)	0.579	2.818	2.591	0.612	2.998	2.989
$\sigma_v$	0.611	2.429	2.484	0.563	2.427	2.403
$\sigma_u$	0.583	2.276	2.508	0.564	2.376	2.629
$\rho$	0.644	–	–	0.647	–	–
<i>n</i> = 200						
<i>g</i> (·)	0.668	2.196	1.988	0.674	2.324	2.437
$\sigma_v$	0.683	2.181	2.108	0.646	2.127	1.996
$\sigma_u$	0.675	2.095	2.152	0.686	1.995	2.173
$\rho$	0.717	–	–	0.718	–	–
<i>n</i> = 400						
<i>g</i> (·)	0.771	1.448	1.409	0.815	1.932	1.437
$\sigma_v$	0.826	1.814	1.778	0.806	1.817	1.766
$\sigma_u$	0.853	1.809	1.745	0.846	1.825	1.754
$\rho$	0.778	–	–	0.784	–	–
<i>n</i> = 800						
<i>g</i>	0.859	1.442	1.823	0.885	1.778	1.884
$\sigma_v$	0.904	1.536	1.936	0.897	1.824	1.747
$\sigma_u$	0.913	1.514	1.942	0.907	1.806	1.679
$\rho$	0.926	–	–	0.921	–	–

ified parametric model but ignoring endogeneity (MLE-C), and the semiparametric model of Fan et al. (1996) that is also ignoring endogeneity (FLW). The main implication of our results is that ignoring endogeneity when it is present can have serious consequences for the estimation of productivity measures and efficiency.

For the smooth-coefficient frontier model, the results in Table 5 show the proposed two-step estimator is well behaved in finite samples in the sense that when the sample size increases, the MSE of the estimated coefficient functions reduces. In addition, when the sample sizes doubled, the MSE of the estimated variance parameters reduces to roughly half indicating  $\sqrt{n}$ -consistency is achieved for these estimated parameters.

## 5 Empirical illustration

### 5.1 Data and model

As an application, we use an input distance function of large US banks based on the same data as in Malikov et al. (2016b), from whom we borrow heavily in this discussion. The data on commercial banks come from Call Reports from the Federal

**Table 4** Ratio of AMSPE and MSE of Each Estimator Relative to CSPLIML and LLIML-DGP 2 ( $\lambda = 2, \rho = 0.5$ )

	Relative to CSPLIML			Relative to LLIML		
	MLE-CE	MLE-C	FLW	MLE-CE	MLE-C	FLW
<i>n</i> = 100						
<i>g</i> (·)	0.588	2.723	2.675	0.601	2.918	2.899
$\sigma_v$	0.621	2.642	2.841	0.613	2.829	2.663
$\sigma_u$	0.576	2.524	2.422	0.581	2.634	2.653
$\rho$	0.643	–	–	0.649	–	–
<i>n</i> = 200						
<i>g</i> (·)	0.682	2.124	2.103	0.696	2.618	2.615
$\sigma_v$	0.704	2.457	2.259	0.692	2.504	2.269
$\sigma_u$	0.681	2.267	2.155	0.699	2.392	2.345
$\rho$	0.738	–	–	0.726	–	–
<i>n</i> = 400						
<i>g</i> (·)	0.852	1.536	1.558	0.778	1.876	1.826
$\sigma_v$	0.846	1.827	1.654	0.768	1.725	1.707
$\sigma_u$	0.897	1.814	1.679	0.735	1.717	1.735
$\rho$	0.826	–	–	0.807	–	–
<i>n</i> = 800						
<i>g</i> (·)	0.913	1.736	1.883	0.842	1.974	1.947
$\sigma_v$	0.902	1.757	1.845	0.831	1.934	1.826
$\sigma_u$	0.947	1.773	1.847	0.827	1.902	1.809
$\rho$	0.882	–	–	0.876	–	–

Reserve Bank of Chicago for 2001:Q1–2010:Q4. We use a selected sub-sample of homogeneous large banks, namely those with total assets in excess of one billion dollars (in 2005 US dollars) in the first 3 years of observation. The sample is an unbalanced panel with 2397 bank–year observations for 285 banks.

We define the following desirable outputs of a bank’s production process: consumer loans ( $y_1$ ), real estate loans ( $y_2$ ), commercial and industrial loans ( $y_3$ ) and securities ( $y_4$ ). We also include off-balance-sheet income ( $y_5$ ) as an additional output. The variable inputs are labor, i.e., the number of full-time equivalent employees ( $x_1$ ), physical capital ( $x_2$ ), purchased funds ( $x_3$ ), interest-bearing transaction accounts ( $x_4$ ) and non-transaction accounts ( $x_5$ ).

Malikov et al. (2016b) also have input price information which we take into account here as instruments in  $Z_{2i}$  (which includes four log relative prices normalized by the price of  $x_1$ ). Moreover, we follow Färe et al. (2005) and standardize all variables prior to the estimation by subtracting their respective sample means and dividing by their sample standard deviations.

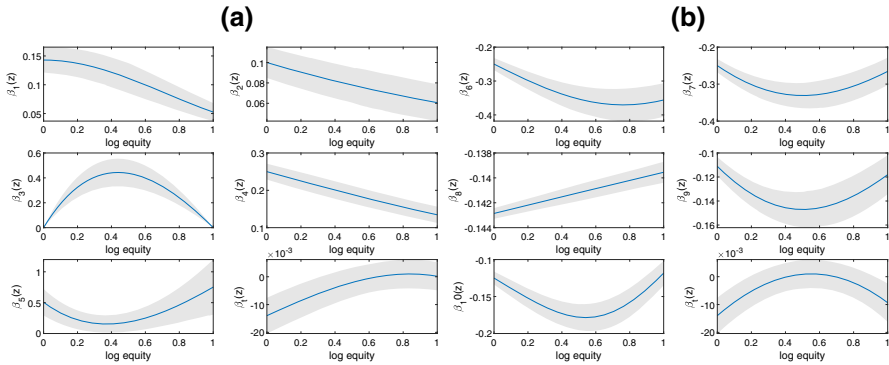
We use the above data to estimate both the nonparametric frontier and the smooth-coefficient frontier models. However, our objective is not to determine and compare the performance of both models but merely to illustrate the usefulness of our proposed

**Table 5** MSE-DGP 3 (Smooth-coefficient frontier)

	$\lambda = 1$			$\lambda = 2$		
	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
<i>n</i> = 100						
$\alpha(\cdot)$	0.115	0.113	0.112	0.092	0.085	0.081
$\beta_1(\cdot)$	0.077	0.072	0.068	0.062	0.053	0.047
$\beta_2(\cdot)$	0.064	0.058	0.049	0.042	0.033	0.029
$\sigma_v$	0.615	0.613	0.588	0.248	0.241	0.239
$\sigma_u$	0.559	0.545	0.546	0.284	0.289	0.277
$\delta$	0.289	0.290	0.281	0.087	0.088	0.087
<i>n</i> = 200						
$\alpha(\cdot)$	0.102	0.099	0.094	0.077	0.073	0.068
$\beta_1(\cdot)$	0.065	0.058	0.047	0.034	0.029	0.024
$\beta_2(\cdot)$	0.055	0.051	0.042	0.031	0.026	0.021
$\sigma_v$	0.312	0.289	0.265	0.143	0.131	0.110
$\sigma_u$	0.284	0.269	0.255	0.144	0.125	0.109
$\delta$	0.140	0.121	0.108	0.044	0.030	0.022
<i>n</i> = 400						
$\alpha(\cdot)$	0.092	0.088	0.082	0.065	0.061	0.058
$\beta_1(\cdot)$	0.045	0.041	0.036	0.023	0.020	0.017
$\beta_2(\cdot)$	0.044	0.040	0.037	0.024	0.022	0.018
$\sigma_v$	0.155	0.143	0.132	0.071	0.065	0.055
$\sigma_u$	0.141	0.135	0.128	0.071	0.063	0.054
$\delta$	0.069	0.060	0.055	0.020	0.014	0.010
<i>n</i> = 800						
$\alpha(\cdot)$	0.084	0.080	0.079	0.062	0.059	0.054
$\beta_1(\cdot)$	0.043	0.037	0.032	0.021	0.018	0.014
$\beta_2(\cdot)$	0.042	0.038	0.033	0.020	0.019	0.015
$\sigma_v$	0.074	0.071	0.065	0.038	0.032	0.027
$\sigma_u$	0.082	0.067	0.061	0.037	0.032	0.029
$\delta$	0.035	0.030	0.024	0.011	0.005	0.004

approaches. Moreover, it is important to note that since we are using panel data, whereas our proposed models are designed for cross-section data, and we need to make assumptions on the temporal behavior of inefficiency,  $u_{it}$  and noise,  $v_{it}$ . For simplicity, we assume both  $u_{it}$  and  $v_{it}$  to be independent and identically distributed for nonparametric specification of the frontier model, and independently distributed for the smoothing coefficient frontier model. In addition, we assume no specific temporal behavior on inefficiency which implies a bank can be fully efficient in one year but not in others.

The first model we consider is the smooth-coefficient input distance function in which the translog functional form is used for the specification of the frontier. The



Note: Log-equity and time are normalized in the interval [0, 1].

**Fig. 1**  $\beta_j(Z)$  as a function of log equity. Note: Log-equity and time are normalized in the interval [0, 1]

translog distance function (IDF) has the form:

$$-\ln x_{Kit} = \beta_0 + \sum_{m=K+1}^{K+M} \beta_m(Z_{it1}) \ln y_m + \sum_{k=1}^K \beta_k(Z_{it1}) \ln (x_{kit}/x_{Kit}) + v_{it}. \quad (22)$$

The variables in  $Z_{it1}$  are a time trend and the log equity capital to account for possible size effects. On the other hand,  $Z_{it2}$  includes  $Z_{it1}$  plus the four log relative prices and non-performing loans in the previous period. We expect  $\beta_m(Z_{it1}) < 0$  ( $m = K + 1, \dots, K + M$ ) and  $\beta_k(Z_{it1}) > 0$  ( $k = 1, \dots, K$ ).

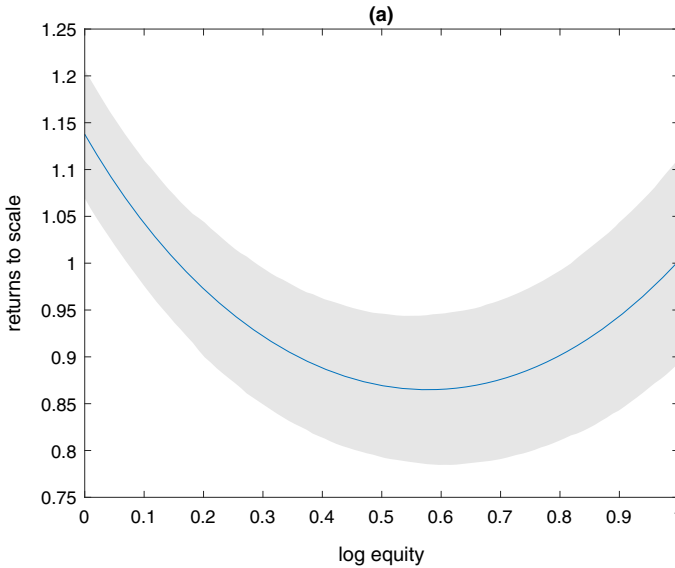
### 5.2 Empirical results

Our empirical results are summarized in Figs. 1, 2 and 3. In Fig. 1, we plot  $\beta_j$ s as a function of log equity ( $Z$ ) which we normalize in the interval [0, 1]. In panel (a) we report the output-related coefficients and in panel (b) the input-related coefficients. As necessitated by the theory, the output-related coefficients are negative and the input-related coefficients are positive. Moreover, the input distance function should be linearly homogeneous with respect to inputs. Technical change is mostly negative except for values of equity near the median, where it also attains its maximum. As a function of equity, these coefficients show a variety of patterns which makes it necessary to use nonparametric estimation.

Next, we report the estimates of returns to scale (RTS) and the results are given in Fig. 2. The RTS can be estimated using the following equation:

$$RTS(z) = - \left( \sum_{m=1}^M \beta_m(z) \right)^{-1}. \quad (23)$$

The results in Fig. 2 indicate that large US banks exhibit, in general, decreasing returns to scale except for relatively small banks that are in, approximately, the 30-th percentile



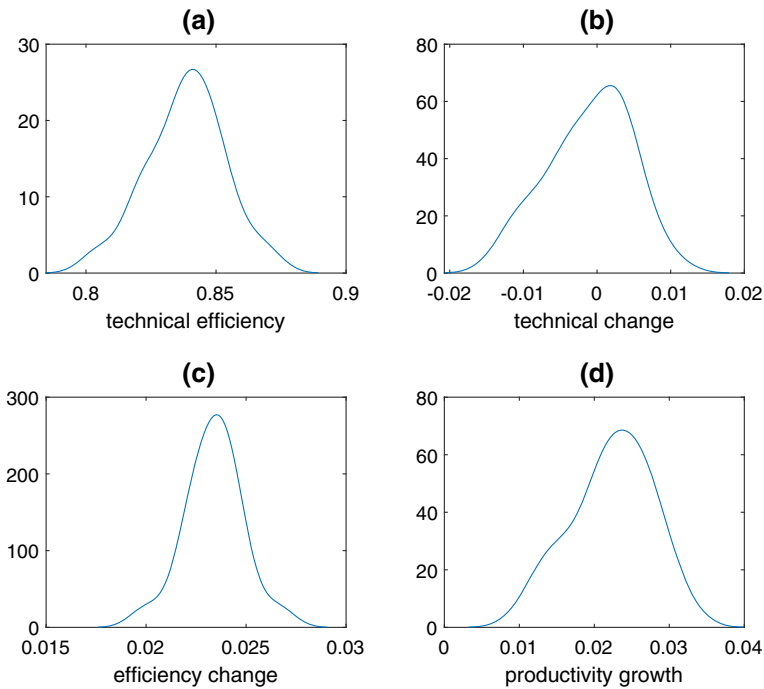
**Fig. 2** Returns to scale

of the log-equity distribution. Returns to scale are lower at the median and are close to one for larger banks; see also the evidence in Restrepo-Tobón and Kumbhakar (2015).

Figure 3 reports the sampling distributions of the estimated technical efficiency (panel (a)) and productivity measures such as technical change (panel (b)), efficiency change (panel (c)) and productivity growth (panel (d)) which is defined as the sum of technical change and efficiency change. The results in Fig. 3 show that both efficiency change and productivity growth are mainly positive, averaging about 2.4% and 2.7%, respectively, while technical change is close to zero on average and extends from roughly -1.8% to 1.6%. The estimated technical efficiency scores are ranging from 74% to 88% with an average of 83% indicating that for some banks, there is more room for improvement.

Next, we consider the estimation of nonparametric input distance frontier model where the frontier is left unspecified. For this model, we use both the local LIML and constrained LIML approaches discussed in Sects. 2.1.1 and 2.1.2. Moreover, for comparison purposes, we also estimate a parametric translog frontier as well as a non-parametric frontier model assuming all regressors are exogenous. These comparisons are particularly important as they highlight the impact of endogeneity. For simplicity and ease of presentation, we summarize the results in the form of sample distributions and these results are given in Fig. 4.

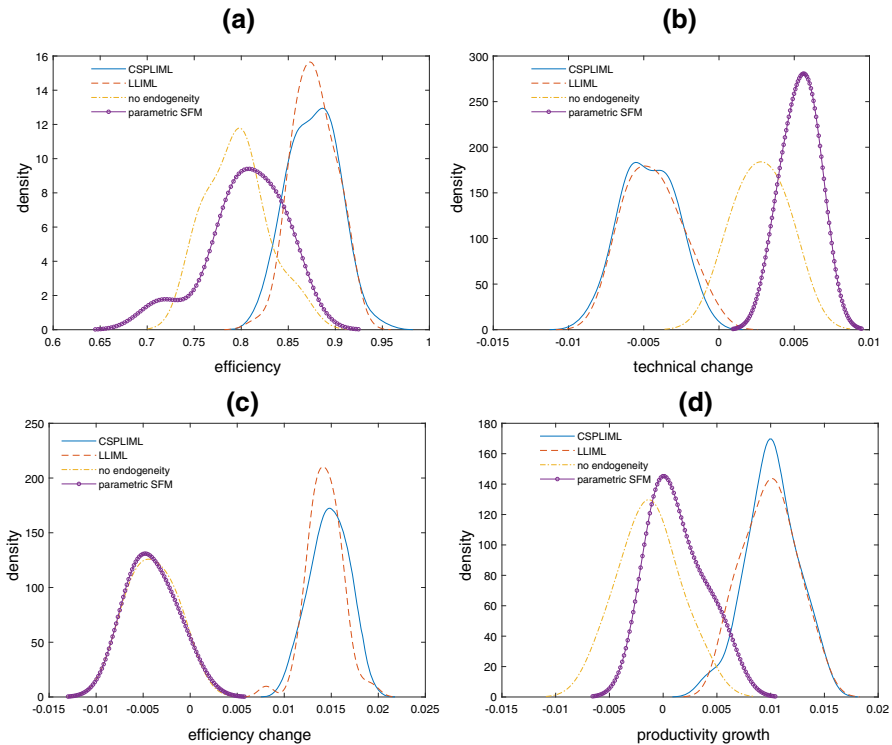
Panel (a) in Fig. 4 reports sample distributions of efficiency obtained using LLIML, CSPLIML, a nonparametric model without endogeneity and a parametric (translog) specification that also does not allow for endogeneity. Efficiency distributions are rather similar for LLIML and CSPLIML, and the same is true for the measures of technical change (panel (b)), efficiency change (panel (c)), and productivity growth (panel (d)). Technical change and efficiency change distributions are quite different



**Fig. 3** Sample distributions of efficiency, efficiency change, technical change and productivity growth: smooth-coefficient model

when endogeneity is ignored. For example, according to local LIML and constrained LIML, technical change averages -0.5% and extends from roughly -1% to slightly over zero, whereas according to the translog and nonparametric model without endogeneity, it averages at 0.5% and spreads from -0.5% to nearly 1%. For efficiency changes, our new techniques give an average close to 1.5%, but the prediction of the models that do not account for endogeneity is that efficiency change averages -0.5% and ranges from roughly -1.5% to slightly over 0.5%. Therefore, there are important differences in terms of productivity growth as local LIML and constrained LIML imply an average value close to 1%, ranging from about zero to nearly 2%. The models that ignore endogeneity imply an average value close to zero, ranging from -1% to 1%, approximately.

The drastic impact of accounting for endogeneity on the average estimated productivity growth in Fig. 4 agrees with what was found in earlier studies. For example, Malikov et al. (2016b) report an average growth rate of 1.1% and a similar sensitivity of estimates to controlling for endogeneity of inputs. They obtain their results using a Bayesian estimation of a directional technology distance function and cost-minimization conditions on the same data set.



**Fig. 4** Sample distributions of efficiency, technical change, efficiency change and productivity growth: nonparametric frontier model

## 6 Concluding remarks

In this paper, we consider the problem of estimating a nonparametric stochastic frontier model with shape restrictions and when some or all regressors are endogenous. The endogeneity is caused by correlation between the symmetric error terms in the production function equation and in the reduced form.

First we propose two estimation procedures that share common features. Both are implemented in a three-step algorithm. Specifically, in the first procedure, we develop a three-step nonparametric constrained limited information maximum likelihood estimator. In the second procedure, we construct a local parametric limited information maximum likelihood estimator, where the constrained nonparametric estimation in the first procedure (Step 2) is replaced by a local parametric approximation.

Additionally, we consider estimation of semiparametric smooth-coefficient stochastic frontier models with endogenous regressors and propose a two-step procedure based on local (or nonparametric) GMM and MLE estimators.

Our Monte Carlo simulations demonstrate an attractive finite sample performance for all estimators proposed in this paper. An empirical application to US banks is presented to illustrate the usefulness of our proposed methods in practice.



Finally, the paper does not pursue several interesting extensions. For example, we do not consider further sources of endogeneity such as dependence between  $u_i$  and  $\xi_i$ ; we do not consider the problem of testing for parametric versus nonparametric specification of the frontier; nor do we consider the problem of testing constancy of parameters in the smooth-coefficient model. These problems merit a separate paper and are left for future work.

**Acknowledgements** We would like to thank the Editors: Subal Kumbhakar, Christopher Parmeter and Amir Malikov, and two anonymous referees to helpful comments and suggestions that led to substantial improvement of the paper. All the remaining errors are our responsibilities. Artem Prokhorov's research for this paper was supported by a grant from the Russian Science Foundation (Project No.20-18-00365).

## References

- Amsler C, Prokhorov A, Schmidt P (2014) Using copulas to model time dependence in stochastic frontier models. *Econ Rev* 33(5–6):497–522
- Amsler C, Prokhorov A, Schmidt P (2016) Endogeneity in stochastic frontier models. *J Econom* 190:280–288
- Amsler C, Prokhorov A, Schmidt P (2017) Endogenous environmental variables in stochastic frontier models. *J Econom* 199:131–140
- Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Econ* 20:325–332
- Caudill SB, Ford JM, Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *J Bus Econ Stat* 13(1):105–111
- Du P, Parmeter CF, Racine JS (2013) Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Stat Sin* 23:1347–1371
- Fan Y, Li Q, Weersink A (1996) Semiparametric estimation of stochastic production frontier models. *J Bus Econ Stat* 14:460–68
- Freyberger J, Horowitz JL (2015) Identification and shape restrictions in nonparametric instrumental variables estimation. *J Econom* 189:41–53
- Färe R, Grosskopf S, Noh D-W, Weber W (2005) Characteristics of a polluting technology: theory and practice. *J Econom* 126:469–492
- Gozalo P, Linton O (2000) Local nonlinear least squares: using parametric information in nonparametric regression. *J Econom* 99:63–106
- Griffiths WE, Hajargasht G (2016) Some models for stochastic frontiers with endogeneity. *J Econom* 190:341–348
- Hall P, Huang L-S (2001) Nonparametric kernel regression subject to monotonicity constraints. *Ann Stat* 624–647
- Henderson DJ, List JA, Millimet DL, Parmeter CF, Price MK (2012) Empirical implementation of nonparametric first-price auction models. *J Econom* 168:17–28
- Henderson DJ, Parmeter CF (2009) Imposing economic constraints in nonparametric regression: survey, implementation, and extension. *Adv Econom* 25:433–69
- Henderson DJ, Parmeter CF (2015) Model averaging over nonparametric estimators. In: *Essays in Honor of Aman Ullah, advances in econometrics*, vol 36
- Huang CJ, Liu J (1994) Estimation of a non-neutral stochastic frontier production function. *J Prod Anal* 5:171–180
- Jondrow J, Lovell CK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econom* 19:233–238
- Karakaplan M, Kutlu L (2015) Handling endogeneity in stochastic frontier analysis. Available at SSRN 2607276
- Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *J Bus Econ Stat* 9(3):279–286
- Kumbhakar SC, Park BU, Simar L, Tsionas EG (2007) Nonparametric stochastic frontiers: a local maximum likelihood approach. *J Econom* 137:1–27

- Kumbhakar SC, Sun K, Zhang R (2016) Semiparametric smooth coefficient estimation of a production system. *Pac Econ Rev* 21:464–482
- Kutlu L (2010) Battese–Coelli estimator with endogenous regressors. *Econ Lett* 109:79–81
- Kutlu L, Tran KC (2019) Heterogeneity and endogeneity in panel stochastic frontier models. In: *Panel data econometrics*. Elsevier, pp 131–146
- Lin X, Carroll RJ (2000) Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J Am Stat Assoc* 95:520–534
- Malikov E, Kumbhakar SC, Sun Y (2016a) Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. *J Econom* 190:233–251
- Malikov E, Kumbhakar SC, Tsionas MG (2016b) A cost system approach to the stochastic directional technology distance function with undesirable outputs: the case of US banks in 2001–2010. *J Appl Econom* 31:1407–1429
- Martins-Filho C, Yao F (2007) Nonparametric frontier estimation via local linear regression. *J Econom* 141:283–319
- Martins-Filho C, Yao F (2009) Nonparametric regression estimation with general parametric error covariance. *J Multivar Anal* 100:309–333
- McManus DA (1994) Making the Cobb–Douglas functional form an efficient nonparametric estimator through localization
- Parmeter C, Racine J (2019) Nonparametric estimation and inference for panel data models
- Parmeter CF, Racine JS (2013) *Smooth constrained frontier analysis*. Springer, New York, pp 463–488
- Restrepo-Tobón D, Kumbhakar SC (2015) Nonparametric estimation of returns to scale using input distance functions: an application to large US banks. *Empir Econ* 48:143–168
- Simar L, Van Keilegom I, Zelenyuk V (2017) Nonparametric least squares methods for stochastic frontier models. *J Prod Anal* 47:189–204
- Su L, Ullah A (2008) Local polynomial estimation of nonparametric simultaneous equations models. *J Econom* 144:193–218
- Su L, Ullah A, Wang Y (2013) Nonparametric regression estimation with general parametric error covariance: a more efficient two-step estimator. *Empir Econ* 45:1009–1024
- Sun K (2015) Constrained nonparametric estimation of input distance function. *J Prod Anal* 43:85–97
- Sun K, Kumbhakar SC (2013) Semiparametric smooth-coefficient stochastic frontier model. *Econ Lett* 120:305–309
- Tran KC, Tsionas EG (2009) Local GMM estimation of semiparametric panel data with smooth coefficient models. *Econom Rev* 29:39–61
- Tran KC, Tsionas EG (2013) GMM estimation of stochastic frontier model with endogenous regressors. *Econ Lett* 118:233–236
- Tran KC, Tsionas EG (2015) Endogeneity in stochastic frontier models: copula approach without external instruments. *Econ Lett* 133:85–88
- Wang H (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *J Prod Anal* 18:241–253

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.