Jin, Fei; Lee, Lung-Fei

Article

Lasso maximum likelihood estimation of parametric models with singular information matrices

Econometrics

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Jin, Fei; Lee, Lung-Fei (2018) : Lasso maximum likelihood estimation of parametric models with singular information matrices, Econometrics, ISSN 2225-1146, MDPI, Basel, Vol. 6, Iss. 1, pp. 1-24, https://doi.org/10.3390/econometrics6010008

econometrics

MDPI

Article

Lasso Maximum Likelihood Estimation of Parametric Models with Singular Information Matrices

Fei Jin [1,2] and Lung-fei Lee [3,*]

[1] School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China; jin.fei@sufe.edu.cn
[2] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China
[3] Department of Economics, The Ohio State University, Columbus, OH 43210, USA
* Correspondence: lee.1777@osu.edu; Tel.: +1-614-247-8481

Received: 1 December 2017; Accepted: 13 February 2018; Published: 22 February 2018

Abstract: An information matrix of a parametric model being singular at a certain true value of a parameter vector is irregular. The maximum likelihood estimator in the irregular case usually has a rate of convergence slower than the $\sqrt{n}$-rate in a regular case. We propose to estimate such models by the adaptive lasso maximum likelihood and propose an information criterion to select the involved tuning parameter. We show that the penalized maximum likelihood estimator has the oracle properties. The method can implement model selection and estimation simultaneously and the estimator always has the usual $\sqrt{n}$-rate of convergence.

Keywords: penalized maximum likelihood; singular information matrix; lasso; oracle properties

JEL Classification: C13; C18; C51; C52

1. Introduction

It has long been noted that some parametric models may have singular information matrices but still be identifiable. For example, Silvey (1959) finds that the score statistic in a single-parameter identifiable model can be zero for all data and Cox and Hinkley (1974) notice that a zero score can arise in the estimation of variance component parameters. Zero or linearly dependent scores imply that information matrices are singular. Other examples include, among others, parametric mixture models that include one homogeneous distribution (Kiefer 1982), simultaneous equations models (Sargan 1983), the sample selection model (Lee and Chesher 1986), the stochastic frontier function model (Lee 1993), and a finite mixture model (Chen 1995).

Some authors have considered the asymptotic distribution of the maximum likelihood estimator (MLE) in some irregular cases with singular information matrices. Cox and Hinkley (1974) show that the asymptotic distribution of the MLE of variance components can be found after a power reparameterization. Lee (1993) derives the asymptotic distribution of the MLE for parameters in a stochastic frontier function model with a singular information matrix by several reparameterizations so that the transformed model has a nonsingular information matrix. Rotnitzky et al. (2000) consider a general parametric model where the information matrix has a rank being one less than the number of parameters, and derive the asymptotic distribution of the MLE by reparameterizations and investigating high order Taylor expansions of the first order conditions. Typically, the MLEs of some components of the parameter vector in the irregular case may have slower than the $\sqrt{n}$-rate of convergence and have non-normal asymptotic distributions, while the MLE in the regular case has the $\sqrt{n}$-rate of convergence and is asymptotically normally distributed. As a result, for inference purposes, one may need to first test whether the parameter vector takes a certain value at which the information matrix is singular.

We consider the case that the irregularity of a singular information matrix occurs when a subvector of the parameter vector takes a specific true value, while the information matrix at any other value is nonsingular. For example, zero true value of a variance parameter in the stochastic frontier function model, and zero true values of a correlation coefficient and coefficients for variables in the selection equation of a sample selection model can lead to singular information matrices (Lee and Chesher 1986). For such a model, if the true value of the subvector is known and imposed in the model, the restricted model will usually have a nonsingular information matrix for the remaining parameters and the MLE has the usual $\sqrt{n}$-rate of convergence. This reminds us of the oracle properties of the lasso in linear regressions, i.e., it may select the correct model with probability approaching one (w.p.a.1.) and the resulting estimator satisfies the properties as if we knew the true model (Fan and Li 2001). In this paper, we propose to estimate an irregular parametric model by a penalized maximum likelihood (PML) which appends a lasso penalty term to the likelihood function. Without loss of generality, we consider the situation when the information matrix is singular at a zero true value θ_{20} of a subvector θ_2 of the parameter vector θ.[1] We expect that a PML with oracle properties for parametric models can avoid the slow rate of convergence and nonstandard asymptotic distribution for the irregular case. We penalize θ_2 using the Euclidean norm as for the group lasso (Yuan and Lin 2006), since the interest is in whether the whole vector θ_2 rather than its individual components are zero. The penalty term is constructed to be adaptive by using an initial consistent estimator as for the adaptive lasso (Zou 2006) and adaptive group lasso (Wang and Leng 2008), so that the PML can have the oracle properties. In the irregular case, the initial estimate used to construct the adaptive penalty term has a slower rate of convergence than that in the literature, but the lasso approach can still be applied if the tuning parameter is properly selected. We prove the oracle properties under regularity conditions. Consequently, the PML can implement model selection and estimation simultaneously. Because the model with $\theta_{20} \neq 0$ and the restricted one with $\theta_{20} = 0$ imposed have nonsingular information matrices, the PML estimator (PMLE) always has the $\sqrt{n}$-rate of convergence and standard asymptotic distributions.

The PML criterion function has a tuning parameter in the penalty term. In asymptotic analysis, the tuning parameter is assumed to have certain order so that the PML can have the oracle properties. In finite samples, the tuning parameter needs to be chosen. For least square shrinkage methods, the generalized cross validation (GCV) and information criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are often used. While the GCV and AIC cannot identify the true model consistently (Wang et al. 2007), the BIC can (Wang and Leng 2007; Wang et al. 2007; Wang et al. 2009). Zhang et al. (2010) propose a general information criterion (GIC) that can nest the AIC and BIC and show its consistency in model selection. Following Zhang et al. (2010), we propose to choose the tuning parameter by minimizing an information criterion. We show that the procedure is consistent in model selection under regularity conditions. Because of the irregularity in the model, the proposed information criterion can be different from the traditional AIC, BIC and GIC.

Jin and Lee (2017) show that, in a matrix exponential spatial specification model, the covariance matrix of the gradient vector for the nonlinear two stage least squares (N2SLS) criterion function can be singular when a subvector of the parameter vector has the true value zero. They consider the penalized lasso N2SLS estimation of the model. This paper generalizes the lasso method to the ML estimation of the several cited models with singular information matrices. For the model in Jin and Lee (2017), the true parameter vector is in the interior of the parameter space. However, for some irregular models cited above, the true parameter vector is on the boundary of the parameter space. We thus consider also the boundary case in this paper.

The PML approach proposed in this paper can be applied to all of the parametric models with singular information matrices mentioned above, e.g., the sample selection model and the stochastic frontier function model. Since the PMLE has the $\sqrt{n}$-rate of convergence for the components which

[1] A model with the new parameter $\eta = \theta_2 - \theta_{20}$ can be considered in the case of a nonzero θ_{20}.

are not super-consistently estimated, we expect the PMLE to outperform the unrestricted MLE in finite samples for such models in the irregular case, e.g., in terms of smaller root mean squared errors and shorter confidence intervals.

The rest of the paper is organized as follows. Section 2 presents the PML estimation procedure for general parametric models with singular information matrices. Section 3 discusses specifically the PMLEs for the sample selection model and stochastic frontier function model. Section 4 reports some Monte Carlo results. Section 5 concludes. In Appendix A, we derive the asymptotic distribution of the MLE of the sample selection model in the irregular case. Proofs are in Appendix B.

2. PMLE for Parametric Models

Let the data $(y_1, \ldots, y_n)$ be i.i.d. with the probability density function (pdf) $f(y;\theta_0)$, a member of the family of pdf's $f(y;\theta)$, $\theta \in \Theta$, if y's are continuous random variables. If y's are discrete, $f(y;\theta)$ will be a probability mass function. Furthermore, if y's are mixed continuous and discrete random variables, $f(y;\theta)$ will be a mixed probability mass and density function. Assumption 1 is a standard condition for the consistency of the MLE (Newey and McFadden 1994).

Assumption 1. *Suppose that y_i, $i = 1, \ldots, n$, are i.i.d. with pdf (or mixed probability mass and density function) $f(y_i;\theta_0)$ and (i) if $\theta \neq \theta_0$ then $f(y_i;\theta) \neq f(y_i;\theta_0)$ with probability one; (ii) $\theta_0 \in \Theta$, which is compact; (iii) $\ln f(y_i;\theta)$ is continuous at each θ with probability one; (iv) $\mathrm{E}[\sup_{\theta\in\Theta} |\ln f(y;\theta)|] < \infty$.*

Rothenberg (1971) shows that, if the information matrix of a parametric model has constant rank in an open neighborhood of the true parameter vector, then local identification of parameters is equivalent to nonsingularity of the information matrix at the true parameter vector. Local identification is necessary but not sufficient for global identification. For the examples in the introduction, the information matrix of a parametric model is singular when the true parameter vector takes certain value, but it is nonsingular at other values. Thus, the result in Rothenberg (1971) does not apply but the parameters may still be identifiable in all cases.

We consider the case that the information matrix of the likelihood function is singular at θ_0, with a subvector θ_{20} of θ_0 being zero. We propose to estimate $\theta = (\theta_1', \theta_2')'$ by maximizing the following penalized likelihood function

$$Q_n(\theta) = [L_n(\theta) - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\theta_2\|] I(\tilde{\theta}_{2n} \neq 0) + L_n(\theta_1, 0) I(\tilde{\theta}_{2n} = 0), \tag{1}$$

where $L_n(\theta) = \frac{1}{n}\sum_{i=1}^n l_i(\theta)$ is the log likelihood function divided by n with $l_i(\theta) = \ln f(y_i;\theta)$, $\lambda_n > 0$ is a tuning parameter, $\mu > 0$ is a constant, $\tilde{\theta}_{2n}$ is an initial consistent estimator of θ_2, which can be the MLE or any other consistent estimator, $\|\cdot\|$ denotes the Euclidean norm and $I(\cdot)$ is the set indicator. The PMLE $\hat{\theta}_n$ maximizes (1).

Assumption 2. $\tilde{\theta}_{2n} = \theta_{20} + o_p(1)$.

The initial estimator $\tilde{\theta}_{2n}$ can be zero in value, especially when θ_{20} is on the boundary of the parameter space, e.g., a zero variance parameter for the stochastic frontier function model in Section 3.2. With a zero value for $\tilde{\theta}_{2n}$, the PMLE of θ_2 in (1) is set to zero and the value of the PMLE equals that of the restricted MLE with the restriction $\theta_2 = 0$ imposed. The tuning parameter λ_n needs to be positive which tends to zero as the sample size increases.

Assumption 3. $\lambda_n > 0$ *and* $\lambda_n = o(1)$.

We have the consistency of $\hat{\theta}_n$ as long as λ_n goes to zero as n goes to infinity in Assumption 3.

Proposition 1. *Under Assumptions 1–3,* $\hat{\theta}_n = \theta_0 + o_p(1)$.

The convergence rate of $\hat{\theta}_n$ can be derived under regularity conditions. Let $\Theta = \Theta_1 \times \Theta_2$, where Θ_1 and Θ_2 are, respectively, the parameter spaces for θ_1 and θ_2. We investigate the case where θ_{20} is on

the boundary as well as the case where θ_{20} is in the interior $\text{int}(\Theta_2)$ of Θ_2. The rest of parameters θ_{10} are always in the interior of Θ_1. The following regularity condition is required.

Assumption 4. *(i)* $\theta_0 = (\theta_{10}', \theta_{20}')' \in \Theta_1 \times \Theta_2$ *which are compact convex subsets in some finite dimensional Euclidean space* $\mathbb{R}^k$*; (ii)* $\theta_{10} \in \text{int}(\Theta_1)$*; (iii)* $\Theta_2 = [0, \zeta)$ *for some* $\zeta > 0$ *if* $\theta_2 \in \mathbb{R}^1$*, and* $\theta_{20} \in \text{int}(\Theta_2)$ *if* $\theta_2 \in \mathbb{R}^{k_2}$ *with* $k_2 \geq 2$*; (iv)* $f(y_i;\theta)$ *is twice continuously differentiable and* $f(y;\theta) > 0$ *on* $\mathcal{S}$*, where* $\mathcal{S} = \mathcal{N}(\theta_0) \cap (\Theta_1 \times \Theta_2)$ *with* $\mathcal{N}(\theta_0)$ *being an open neighborhood at* θ_0 *of* $\mathbb{R}^k$*; (v)* $\int \sup_{\theta \in \mathcal{S}} \| \frac{\partial f(y;\theta)}{\partial \theta} \| \, dy < \infty$*,* $\int \sup_{\theta \in \mathcal{S}} \| \frac{\partial^2 f(y;\theta)}{\partial \theta \partial \theta'} \| \, dy < \infty$*; (vi)* $\text{E}(\frac{\partial l_i(\theta_0)}{\partial \theta} \frac{\partial l_i(\theta_0)}{\partial \theta'})$ *exists and is nonsingular when* $\theta_{20} \neq 0$*, and* $\text{E}(\frac{\partial l_i(\theta_0)}{\partial \theta_1} \frac{\partial l_i(\theta_0)}{\partial \theta_1'})$ *exists and is nonsingular when* $\theta_{20} = 0$*; (vii)* $\text{E}(\sup_{\theta \in \mathcal{S}} \| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta'} \|) < \infty$.

In the literature, several irregular models have parameters on the boundary: the model on simplified components of variances in Cox and Hinkley (1974, p. 117), the mixture model in Kiefer (1982) and the stochastic frontier function model in Aigner et al. (1977).[2] For these models, a scalar parameter θ_2 is always nonnegative but irregularity occurs when $\theta_{20} = 0$ on the boundary. True parameters other than θ_{20} are in the interior of their parameter spaces. We thus assume that θ_{20} is a scalar when it can be on the boundary of its parameter space.[3] (iv)–(vii) in Assumption 4 are standard. Note that for the partial derivative with respect to θ_2 at θ_{20} on the boundary, only perturbations on Θ_2 are considered, as for the (left/right) partial derivatives in Andrews (1999). The convexity of Θ_1 and Θ_2 makes such derivatives well-defined and convexity is relevant when the mean value theorem is applied to the log likelihood function.

For our main focus in this paper, at $\theta_{20} = 0$, the information matrix is singular. However, our lasso estimation method is also applicable to regular models where the information matrix might be nonsingular even at $\theta_{20} = 0$. The following proposition provides such a generality.

Proposition 2. *Under Assumptions 1–4, if* $\text{E}(\frac{\partial l_i(\theta_0)}{\partial \theta} \frac{\partial l_i(\theta_0)}{\partial \theta'})$ *exists and is nonsingular, then* $\hat{\theta}_n = \theta_0 + O_p(n^{-1/2} + \lambda_n)$.

Proposition 2 derives the rate of convergence of the PMLE $\hat{\theta}_n$ in the case of a nonsingular information matrix. When $\theta_{20} \neq 0$, we have assumed in Assumption 4 that the information matrix is nonsingular. When $\theta_{20} = 0$, Proposition 2 is relevant in the event that the PML is formulated with a reparameterized model that has a nonsingular information matrix and the reparameterized unknown parameters are represented by θ.

We now consider whether the PMLE has the sparsity property, i.e., whether $\hat{\theta}_{2n}$ is equal to zero w.p.a.1. when $\theta_{20} = 0$. For the lasso penalty function, λ_n and the initial consistent estimate $\tilde{\theta}_n$ are required to have certain orders of convergence for the sparsity property.

Assumption 5. *Suppose that* $\tilde{\theta}_n - \theta_0 = O_p(n^{-s})$*, where* $0 < s \leq 1/2$*. The tuning parameter sequence* λ_n *is selected to satisfy either*

(i) λ_n *converges to zero such that* $\lambda_n n^{\mu s} \to \infty$ *as* $n \to \infty$*; or*
(ii) if $\text{E}(\frac{\partial l_i(\theta_0)}{\partial \theta} \frac{\partial l_i(\theta_0)}{\partial \theta'})$ *exists and is nonsingular,* λ_n *is selected to have at most the order* $O(n^{-1/2})$ *such that* $\lambda_n n^{\mu s + 1/2} \to \infty$ *as* $n \to \infty$.

---

[2] As pointed out by an anonymous referee, our PML approach can also be applied to interesting economic models such as disequilibrium models and structural change models. For a market possibly in disequilibrium, an equilibrium is characterized by a parameter value on the boundary (Goldfeld and Quandt 1975; Quandt 1978). Structural changes can also be characterized by parameters on the boundary. Thus, our PML approach can be applied in those models with singular information matrices.

[3] This implies that $\theta_0 \in \text{int}(\Theta)$ when $\theta_{20} \neq 0$, which simplifies later presentation for the asymptotic distribution of the PMLE. In the case that $\theta_2 \in \mathbb{R}^{k_2}$ with $k_2 \geq 2$ and θ_{20} is allowed to be on the boundary of Θ_2, when $\theta_{20} \neq 0$, some components of θ_{20} can still be on the boundaries of their parameter spaces, then the asymptotic distributions of their PMLEs will be nonstandard.

According to Rotnitzky et al. (2000), in the case that the information matrix is singular with rank being one less than the number of parameters k, there exists a reparameterization such that the MLE of one of the transformed parameter component converges at a rate slower than $\sqrt{n}$, but the remaining $k-1$ transformed components converge at the $\sqrt{n}$-rate. As a result, some components of the MLE in terms of the original parameter vector have a slower than the $\sqrt{n}$-rate of convergence, while the remaining components may have the $\sqrt{n}$-rate. In this case, for $\tilde{\theta}_n$ as a whole, $s < 1/2$ in Assumption 5 if $\tilde{\theta}_n$ is the MLE. Assumption 5 (i) can be satisfied if λ_n is selected to have a relatively slow rate of convergence to 0. The condition differs from that in the literature due to the irregularity issue we are considering. In the case that the PML is formulated with a reparameterized model that has a nonsingular information matrix and θ represents the reparameterized unknown parameter vector, Assumption 5 (ii) is relevant with $s = 1/2$ if $\tilde{\theta}_n$ is the MLE.

The oracle properties of the PMLE, including the sparsity property, are presented in Proposition 3.[4] When $\theta_{20} = 0$, the PMLE $\hat{\theta}_{2n}$ of θ_2 can equal zero w.p.a.1., and $\hat{\theta}_{1n}$ has the same asymptotic distribution as that of the MLE as if we knew $\theta_{20} = 0$.

Proposition 3. *Under Assumptions 1–5, if* $\theta_{20} = 0$, *then* $\lim_{n\to\infty} \mathrm{P}(\hat{\theta}_{2n} = 0) = 1$, *and* $\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \xrightarrow{d} N\big(0, (-\mathrm{E}\,\frac{\partial^2 l(\theta_0)}{\partial\theta_1 \partial\theta_1'})^{-1}\big)$.

We next turn to the case with $\theta_{20} \neq 0$. The consistency of $\hat{\theta}_n$ to θ_0 in Proposition 1 will guarantee that $P(\hat{\theta}_{2n} \neq 0)$ goes to 1 if $\theta_{20} \neq 0$. By Proposition 2, in order that $\hat{\theta}_n$ can converge to θ_0 with $\sqrt{n}$-consistency and without an asymptotic impact of the first order by λ_n when $\theta_{20} \neq 0$, we need to select λ_n to converge to zero with the order $o(n^{-1/2})$.

Assumption 6. $\lambda_n = o(n^{-1/2})$.

Assumptions 5 and 6 need to coordinate with each other as they are opposite requirements. By taking $\lambda_n = O(n^{-\tau})$ for some $\tau > 1/2$, Assumption 6 holds. Assumption 5 (i) can be satisfied if we take μ to be large enough such that $\mu s > \tau > 1/2$. For such a τ to exist, it is necessary to take $\mu > 1/(2s)$ for a given s. For the regular case in Assumption 5 (ii) , it is relatively more flexible on the value of μ.

Proposition 4. *Under Assumptions 1–4 and 6, if* $\theta_{20} \neq 0$, $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$. *Furthermore, as* $\theta_0 \in \mathrm{int}(\Theta)$, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (-\mathrm{E}\,\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'})^{-1})$.

We next consider the selection of the tuning parameter λ_n. To make explicit the dependence of the PMLE $\hat{\theta}_n$ on a tuning parameter λ, denote the PMLE $\hat{\theta}_\lambda = \arg\max_{\theta\in\Theta}\{[L_n(\theta) - \lambda\|\tilde{\theta}_{2n}\|^{-\mu}\|\theta_2\|]I(\tilde{\theta}_{2n} \neq 0) + L_n(\theta_1, 0)I(\tilde{\theta}_{2n} = 0)\}$ for a given λ.[5] Let $\Lambda = [0, \lambda_{\max}]$ be an interval from which the tuning parameter λ is selected, where $\lambda_{\max}$ is a finite positive number. We propose to select the tuning parameter that maximizes the following information criterion:

$$H_n(\lambda) = L_n(\hat{\theta}_\lambda) + \Gamma_n I(\hat{\theta}_{2\lambda} = 0), \tag{2}$$

where $\{\Gamma_n\}$ is a positive sequence of constants, and $\hat{\theta}_{2\lambda}$ is the PMLE of θ_2 for a given λ. That is, given Γ_n, the selected tuning parameter is $\hat{\lambda}_n = \arg\max_{\lambda\in\Lambda} H_n(\lambda)$. The term Γ_n is an extra bonus for setting θ_2 to zero. Some conditions on Γ_n are also invoked.

Assumption 7. $\Gamma_n > 0$, $\Gamma_n \to 0$ *and* $n^{2s}\Gamma_n \to \infty$ *as* $n \to \infty$.

---

[4] Proposition 2 is proved in the case of a nonsingular information matrix, similar to that in Fan and Li (2001). The method cannot be used in the case of a singular information matrix. However, the sparsity property can still be established by using only the consistency of $\hat{\theta}_n$ under Assumption 5 (i).

[5] As before, when $\tilde{\theta}_{2n} = 0$, the PMLE of θ_2 is $\hat{\theta}_{2\lambda} = 0$.

To balance the order requirements of $\Gamma_n \to 0$ and $n^{2s}\Gamma_n \to \infty$, Γ_n can be taken to be $O(n^{-s})$. As this order changes with s, the information criterion in (2) can be different from the traditional ones such as the AIC, BIC and Hannan-Quinn information criterion.

Let $\{\bar{\lambda}_n\}$ be an arbitrary sequence of tuning parameters which satisfy Assumptions 3, 5 and 6, e.g., $\bar{\lambda}_n = n^{-(\mu s)/2-1/4}$, where μ is chosen such that $\mu s > 1/2$. Define $\Lambda_n = \{\lambda \in \Lambda : \hat{\theta}_{2\lambda} = 0 \text{ if } \theta_{20} \neq 0, \text{ and } \hat{\theta}_{2\lambda} \neq 0 \text{ if } \theta_{20} = 0\}$. In Proposition 5, we let the initial estimator $\tilde{\theta}_n$ be the MLE.

Proposition 5. *Under Assumptions 1–7,* $\mathrm{P}(\sup_{\lambda\in\Lambda_n} H_n(\lambda) < H_n(\bar{\lambda}_n)) \to 1$ *as* $n \to \infty$.

Proposition 5 states that the model selection by the tuning parameter selection procedure is consistent. It implies that any λ in Λ_n that fails to identify the true model would not be selected asymptotically by the information criterion in (2) as an optimal tuning parameter in Λ_n, because such a λ is less favorable than any $\bar{\lambda}_n$, which can identify the true model asymptotically.

3. Examples

In this section, we illustrate the PMLEs of the sample selection model as well as the stochastic frontier function model. In the irregular case, the true parameter vector is in the interior of its parameter space for the sample selection model, but it is on the boundary for the stochastic frontier function model.

3.1. The Sample Selection Model

We consider the sample selection model in Lee and Chesher (1986), which can have a singular information matrix. The model is as follows:

$$y_i = x_i'\beta + \epsilon_i, \quad y_i^* = z_i'\gamma - u_i, \quad i = 1, \ldots, n, \tag{3}$$

where n is the sample size, (x_i, z_i) is the ith observation of exogenous variables, and the vectors (ϵ_i, u_i), for $i = 1, \ldots, n$, are independently distributed as the bivariate normal $N\left(0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right)$. The variable y_i^* is not observed, but a binary indicator I_i is observed to be 1 if and only if $y_i^* \geq 0$ and I_i is 0 otherwise. The variable y_i is only observed when $I_i = 1$. Let $\theta = (\beta', \sigma^2, \gamma', \rho)'$, $\beta = (\beta_1, \beta_2')'$ and $\gamma = (\gamma_1, \gamma_2')'$, where β_1 and γ_1 are, respectively, the coefficients for the intercept terms in the outcome and selection equations. According to Lee and Chesher (1986), when x_i contains an intercept term, but the true values of γ_2 and the correlation coefficient ρ are zero, elements of the score vector are linearly dependent and the information matrix is singular.[6] For this model, the true parameter vector θ_0 which causes irregularity is in the interior of the parameter space.

We derive the asymptotic distribution of the MLE in this irregular case in Appendix A.[7] Let $\tilde{\theta}_n$ be the MLE of θ. It is shown that, for $(\gamma_{20}', \rho_0)' \neq 0$, all components of $\tilde{\theta}_n$ have the usual $\sqrt{n}$-rate of convergence and are asymptotically normal. However, at $(\gamma_{20}', \rho_0)' = 0$, $n^{1/6}\tilde{\rho}_n$ has the same asymptotic distribution as that of $(n^{1/2}\tilde{r}_n)^{1/3}$, where $\tilde{r}_n$ is a transformed parameter and $n^{1/2}\tilde{r}_n$ is asymptotically normal, $n^{1/6}(\tilde{\beta}_{1n} - \beta_{10})$ has the same asymptotic distribution as that of $\sigma_0\psi_0(n^{1/2}\tilde{r}_n)^{1/3}$, where $\psi_0 = \phi(\gamma_{10})/\Phi(\gamma_{10})$, and $n^{1/3}(\tilde{\sigma}_n^2 - \sigma_0^2)$ has the same asymptotic distribution as that of $\sigma_0^2\psi_0(\psi_0 + \gamma_{10})(n^{1/2}\tilde{r}_n)^{2/3}$, while $n^{1/2}(\tilde{\beta}_{2n}' - \beta_{20}', \tilde{\gamma}_n' - \gamma_0')'$ is asymptotically normal. Thus, $\tilde{\rho}_n$, $\tilde{\beta}_{1n}$ and $\tilde{\sigma}_n^2$ have slower than the $\sqrt{n}$-rate of convergence, but $\tilde{\beta}_{2n}$ and $\tilde{\gamma}_n$ have the usual $\sqrt{n}$-rate of convergence.

---

[6] Another irregular case is that z_i consists of only a constant term and dichotomous explanatory variables, and x_i contains the same set of dichotomous explanatory variables and their interaction terms. For this case, the reparameterization process discussed in Appendix A to derive the asymptotic distribution of the MLE also applies.

[7] The method is similar to that in Lee (1993) for the stochastic frontier function model.

Let $\theta_1 = (\beta', \sigma^2, \gamma_1)'$ and $\theta_2 = (\gamma_2', \rho)'$. The PML criterion function for model (3) with the MLE $\tilde{\theta}_{2n}$ is

$$[L_n(\theta) - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\theta_2\|] I(\tilde{\theta}_{2n} \neq 0) + L_n(\theta_1, 0) I(\tilde{\theta}_{2n} = 0). \tag{4}$$

Since $\tilde{\gamma}_{2n} = O_p(n^{-1/2})$ and $\tilde{\rho}_n = O_p(n^{-1/6})$, Assumptions 5 (i) and 6 hold when μ is greater than 3. By Assumption 7, in the information criterion function (2), Γ_n should satisfy $\Gamma_n \to 0$ and $n^{1/3}\Gamma_n \to \infty$ as $n \to \infty$.

According to the discussions in deriving the asymptotic distribution of the MLE via reparameterizations, alternatively, the criterion function for the PMLE can be formulated with the function $L_{n3}(\eta, r)$ of the transformed parameters as

$$\begin{aligned} &[L_{n3}(\eta, r) - \lambda_n \|\tilde{\omega}_n\|^{-\mu_1} \|\omega\|] I(\tilde{\omega}_n \neq 0) + L_{n3}(\eta_1, 0) I(\tilde{\omega}_n = 0) \\ &= [L_n(\theta) - \lambda_n \|\tilde{\omega}_n\|^{-\mu_1} \|\omega\|] I(\tilde{\theta}_{2n} \neq 0) + L_n(\theta_1, 0) I(\tilde{\theta}_{2n} = 0). \end{aligned} \tag{5}$$

where $\eta = (\beta_1 - \sigma_0\lambda_0\rho, \beta_2', \sigma^2 - \rho^2\sigma_0^2\lambda_0(\lambda_0 + \gamma_{10}), \gamma')' = (\varpi', \gamma_2')'$, $r = \rho^3$, and $\omega = (\gamma_2', r)'$. While γ_2 enters the penalty terms of (4) and (5) in the same way, it is not the case for ρ: it is ρ in (4) but ρ^3 in (5). Since $L_{n3}(\eta, r)$ has a nonsingular information matrix, by Proposition 2, the PMLE has the order $O_p(n^{-1/2} + \lambda_n)$, which is $O_p(n^{-1/2})$ under the assumption $\lambda_n = o(n^{-1/2})$. Then $\lambda_n n^{\mu_1 s + 1/2} = \lambda_n n^{(\mu_1+1)/2} \to \infty$ as $n \to \infty$ in Assumption 5 (ii) will be relevant. Thus, for the PML criterion function (5), as long as $\mu_1 > 0$, no further condition on μ_1 is needed. Furthermore, Assumption 7 for Γ_n in the information criterion function (2) with $\Gamma_n \to 0$ and $n\Gamma_n \to \infty$ as $n \to \infty$ is relevant, and we can take $\Gamma_n = O(n^{-1/2})$.

3.2. The Stochastic Frontier Function Model

Consider the following stochastic frontier function model:

$$y_i = x_i'\beta + u_i + v_i, i = 1, \ldots, n, \tag{6}$$

where x_i is a k-dimensional vector of exogenous variables which contains a constant term, the disturbance $u_i \leq 0$ represents technical inefficiency, v_i represents uncontrollable disturbance, and u_i and v_i are independent. Following the literature, u_i is assumed to be half normal with the pdf

$$h(u) = \frac{2}{\sqrt{2\pi}\sigma_1} \exp(-\frac{u^2}{2\sigma_1^2}), u \leq 0,$$

and $v_i \sim N(0, \sigma_2^2)$. As in Aigner et al. (1977), let $\delta = \sigma_1/\sigma_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. For a random sample of size n, the log likelihood function divided by n is

$$L_n(\theta) = \ln(2) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2n\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \frac{1}{n}\sum_{i=1}^{n}\ln\big[1 - \Phi(\frac{\delta(y_i - x_i'\beta)}{\sigma})\big], \tag{7}$$

where $\theta = (\beta', \sigma^2, \delta)'$. In this model, δ is nonnegative and, for the irregular case, the true parameter $\delta_0 = 0$ lies on the boundary, which represents the absence of technical inefficiency. According to Lee (1993), when $\delta_0 = 0$, the information matrix is singular and the MLE of δ has the convergence rate $n^{-1/6}$; when $\delta_0 \neq 0$, the information matrix has full rank and the MLE has the $\sqrt{n}$-rate of convergence. The asymptotic distribution of the MLE when $\delta_0 = 0$ is derived by transforming the model into one with a nonsingular information matrix via several reparameterizations. Thus, the PML estimation can be formulated similarly to the sample selection model, using the original model or the transformed model. Note that in finite samples, the MLE of δ, regardless of whether $\delta_0 = 0$ or not, can be zero with a positive probability. A necessary and sufficient condition for the MLE of δ to be zero is $\sum_{i=1}^{n}\hat{\epsilon}_i^2 \geq 0$, where $\hat{\epsilon}_i$'s are the least squares residuals (Lee 1993).

4. Monte Carlo

In this section, we report results from some Monte Carlo experiments for both the sample selection model and the stochastic frontier function model. The code files are written and run in MATLAB.

4.1. The Sample Selection Model

For the sample selection model, in the experiments, there are two exogenous variables in x_i: one is an intercept term and the other is drawn randomly from the standard normal distribution. The true vector of coefficients for x_i is $(1, 1)'$. There are also two exogenous variables in z_i: an intercept term with true coefficient 1 and a variable randomly drawn from the standard normal distribution, for which the true coefficient is 2, 0.5 or 0. Two values of σ_0^2, 2 and 0.5, are considered. The ρ_0 is either 0.7, -0.7, 0.3, -0.3 or 0. In the information criterion function (2) for the tuning parameter selection, μ is set to 4 and $\Gamma_n = 0.26n^{-1/2}$.[8] An estimate is regarded as zero if it is smaller than 10^{-5}. The number of Monte Carlo repetitions is 1000. The sample sizes considered are $n = 200$ or 600.

Table 1 reports the probabilities that the PMLEs select the right model, i.e., the probabilities of the PMLEs of θ_2 being zero when $\theta_{20} = 0$, and being nonzero when $\theta_{20} \neq 0$. We use PMLE-o and PMLE-t to denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. When $\gamma_{20} = 2$ or 0.5, with the sample size $n = 200$, the probabilities are 1 or very closed to 1; with the sample size $n = 600$, all probabilities are 1. When $\gamma_{20} = 0$ and $\rho_0 = 0$, the PMLEs estimate $\theta_2 = (\gamma_2, \rho)'$ as zero with high probabilities, higher than 95% for the PMLE-o and higher than 69% for the PMLE-t. The PMLE-o has higher probabilities of estimating θ_2 as zero than the PMLE-t. As the sample size increases from 200 to 600, the correct model selection probabilities of the PMLE-o increase while those of the PMLE-t decrease. When $\gamma_{20} = 0$ but $\rho_0 \neq 0$, the PMLEs estimate θ_2 as nonzero with very low probabilities. With $\gamma_{20} = 0$, we see that $\psi_0\sigma_0\frac{\partial L_n(\alpha_0,\rho)}{\partial \beta_1} + 2\rho\sigma_0^2\psi_0(\psi_0 + \gamma_{10})\frac{\partial L_n(\alpha_0,\rho)}{\partial \sigma^2} + \frac{\partial L_n(\alpha_0,\rho)}{\partial \rho} = O(\rho^2)$. Thus, the scores are approximately linearly dependent as $|\rho| < 1$. In finite samples, even though $\rho_0 \neq 0$, the identification can be weak and the MLE behaves similarly to that in the case with $\rho_0 = 0$, which has large bias and variance, as seen from Tables 4 and 5 below. As a result, the PMLEs which use the MLEs to construct the penalty terms have low probabilities of estimating θ_2 to be non-zero.

Table 2 presents the biases, standard errors (SE) and root mean squared errors (RMSE) of the estimates when $\gamma_{20} = 2$. For a nonzero true parameter value, the biases, SEs and RMSEs are divided by the absolute value of the true parameter value. The upper panel is for the sample with size $n = 200$. The restricted MLE, denoted as MLE-r, usually has the largest bias, because it imposes the wrong restriction $\theta_2 = 0$. The MLE, PMLE-o and PMLE-t almost have identical summary statistics. Their biases and SEs are relatively low, e.g., the biases of ρ are all below or equal to 0.012, or 2.5% for a nonzero true ρ_0, and the SEs are all below or equal to 0.246. As the SEs dominate the biases, the RMSEs have similar magnitudes as those of the SEs. As the value of ρ_0 changes, the biases, SEs and RMSEs do not change much. When σ_0^2 decreases from 2 to 0.5, all estimates of β_1, β_2 and σ^2 tend to have smaller biases and SEs, but those for γ_1, γ_2 and ρ show little changes. As the sample size increases to 600, all estimates have smaller biases, SEs and RMSEs.

---

[8] In theory, the information criterion (2) can achieve model selection consistency as long as Γ_n satisfies the order requirement in Assumption 7. However, the finite sample performance depends on the choice of Γ_n. From the proof of Proposition 5, when $\theta_{20} \neq 0$, for large enough n, Γ_n should be smaller than the difference between the function values of the expected log density at the true parameter vector and at the probability limit of the restricted MLE with the restriction $\theta_2 = 0$ imposed. When $\theta_{20} = 0$, Γ_n should be larger than the difference of the function values of the likelihood divided by n at the MLE and at the restricted MLE. For $\theta_{20} = 0$, $\sigma_0^2 = 2$ and $n = 200$, we compute the second difference 1000 times, and set $\Gamma_n = kn^{-1/2}$ to be the sample mean plus 2 times the standard error, which yields $k = 0.26$. We then set $\Gamma_n = 0.26n^{-1/2}$ in all cases and for all sample sizes. We also tried setting $\Gamma_n = kn^{-1/2}$ to be the sample mean plus zero to four times the standard error. The results are relatively sensitive to the choice of k. We leave the theoretical study on the choice of the constant in Γ_n to future research.

Table 3 illustrates the biases, SEs and RMSEs of the estimates when $\gamma_{20} = 0.5$. The patterns are similar to those for Table 2. With a smaller γ_0, the biases and SEs of β_2, γ_1 and γ_2 tend to be smaller, but those of β_1, σ^2 and ρ are larger.

Table 4 reports the biases, SEs and RMSEs when $\gamma_{20} = 0$ but $\rho_0 \neq 0$. We observe that the MLE has relatively large biases and SEs. For $n = 200$, the biases of ρ can be as high as 0.46 in absolute value, or higher than 100%, and the SEs can be as high as 0.72. While the biases of the MLE are usually smaller than those of the MLE-r, the SEs are usually much larger, especially for β_1, σ^2 and ρ. In terms of the RMSEs, the MLE does not show an advantage over the MLE-r. The biases of the PMLE-o are usually smaller than those of the MLE-r and larger than those of the MLE, but the SEs of the PMLE-o are generally smaller than those of the MLE. The PMLE-t has smaller biases than those of PMLE-o but larger SEs in most cases, more similar to the MLE. That is consistent with Table 1, since the PMLE-t estimates θ_2 as nonzero with higher probabilities. The RMSEs of the PMLEs are usually smaller than those of the MLE but larger than those of the MLE-r. In this case, even though the PML methods do not provide good probabilities of selecting the non-zero models, the shrinkage feature of the lasso does provide smaller RMSEs than those of the unconstrained MLEs.

The results for $\gamma_{20} = 0$ and $\rho_0 = 0$ are reported in Table 5. As expected, the MLE-r usually has the smallest biases, SEs and RMSEs, since it has imposed the correct restriction $\theta_2 = 0$. The biases, SEs and RMSEs of the PMLEs are between those of the MLE-r and MLE. The PMLE-o of β_1, σ^2, γ_2 and ρ have significantly smaller biases, SEs and RMSEs than those of the MLE. The biases, SEs and RMSEs of the PMLE-t are smaller than those of the MLE, but larger than those of the PMLE-o, since it estimates θ_2 as nonzero with higher probabilities. Note that the MLEs of β_1, σ^2 and ρ have relatively very large SEs, and the MLEs of σ^2 have very large biases, which can be larger than 50%. With a smaller σ_0^2, the estimates generally have smaller biases, SEs and RMSEs. As n increases to 600, the summary statistics of the PMLE-o become very similar to those of the MLE-r, and all estimates have smaller biases, SEs and RMSEs in general.

Table 1. Probabilities that the PMLEs of the sample selection model select the right model.

	$\gamma_{20} = 2$		$\gamma_{20} = 0.5$		$\gamma_{20} = 0$	
	PMLE-o	**PMLE-t**	**PMLE-o**	**PMLE-t**	**PMLE-o**	**PMLE-t**
$n = 200$						
$\sigma_0^2 = 2, \rho_0 = 0.7$	1.000	1.000	0.999	0.999	0.058	0.222
$\sigma_0^2 = 2, \rho_0 = -0.7$	1.000	1.000	1.000	1.000	0.072	0.241
$\sigma_0^2 = 2, \rho_0 = 0.3$	1.000	1.000	0.999	0.999	0.045	0.196
$\sigma_0^2 = 2, \rho_0 = -0.3$	1.000	1.000	0.997	0.999	0.043	0.200
$\sigma_0^2 = 2, \rho_0 = 0$	1.000	1.000	0.999	0.999	0.955	0.808
$\sigma_0^2 = 0.5, \rho_0 = 0.7$	1.000	1.000	1.000	1.000	0.051	0.191
$\sigma_0^2 = 0.5, \rho_0 = -0.7$	1.000	1.000	1.000	1.000	0.050	0.209
$\sigma_0^2 = 0.5, \rho_0 = 0.3$	1.000	1.000	0.998	1.000	0.054	0.216
$\sigma_0^2 = 0.5, \rho_0 = -0.3$	1.000	1.000	0.997	0.997	0.035	0.166
$\sigma_0^2 = 0.5, \rho_0 = 0$	1.000	1.000	0.996	0.998	0.964	0.809
$n = 600$						
$\sigma_0^2 = 2, \rho_0 = 0.7$	1.000	1.000	1.000	1.000	0.014	0.310
$\sigma_0^2 = 2, \rho_0 = -0.7$	1.000	1.000	1.000	1.000	0.007	0.333
$\sigma_0^2 = 2, \rho_0 = 0.3$	1.000	1.000	1.000	1.000	0.004	0.255
$\sigma_0^2 = 2, \rho_0 = -0.3$	1.000	1.000	1.000	1.000	0.003	0.273
$\sigma_0^2 = 2, \rho_0 = 0$	1.000	1.000	1.000	1.000	0.996	0.692
$\sigma_0^2 = 0.5, \rho_0 = 0.7$	1.000	1.000	1.000	1.000	0.008	0.275
$\sigma_0^2 = 0.5, \rho_0 = -0.7$	1.000	1.000	1.000	1.000	0.011	0.244
$\sigma_0^2 = 0.5, \rho_0 = 0.3$	1.000	1.000	1.000	1.000	0.002	0.228
$\sigma_0^2 = 0.5, \rho_0 = -0.3$	1.000	1.000	1.000	1.000	0.001	0.214
$\sigma_0^2 = 0.5, \rho_0 = 0$	1.000	1.000	1.000	1.000	0.997	0.755

The penalized maximum likelihood (PMLE)-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. When $\theta_{20} \neq 0$, the numbers in the table are the probabilities that the PMLEs of θ_2 are non-zero; when $\theta_{20} = 0$, the numbers are the probabilities that the PMLEs of θ_2 are zero.

Table 2. The biases, standard errors (SE) and root mean squared errors (RMSE) of the estimators when $\gamma_{20} = 2$ in the sample selection model.

n, σ_0^2, ρ_0		β_1	β_2	σ^2	γ_1	γ_2	ρ
200, 2, 0.7	MLE-r	−0.344[0.134]0.369	−0.003[0.135]0.135	−0.163[0.260]0.307	−0.001[0.091]0.091	−2.000[0.000]2.000	−0.700[0.000]0.700
	MLE	0.011[0.164]0.164	−0.003[0.129]0.129	−0.022[0.301]0.302	−0.004[0.132]0.132	0.054[0.269]0.274	0.002[0.146]0.146
	PMLE-o	0.011[0.164]0.165	−0.003[0.129]0.129	−0.022[0.302]0.302	−0.004[0.132]0.132	0.053[0.269]0.274	0.003[0.146]0.146
	PMLE-t	0.011[0.164]0.164	−0.003[0.129]0.129	−0.022[0.301]0.302	−0.004[0.132]0.132	0.054[0.269]0.274	0.002[0.146]0.146
200, 2, −0.7	MLE-r	0.359[0.140]0.385	0.000[0.138]0.138	−0.161[0.252]0.299	−0.002[0.088]0.088	−2.000[0.000]2.000	0.700[0.000]0.700
	MLE	0.001[0.171]0.171	0.000[0.130]0.130	−0.017[0.296]0.296	−0.001[0.134]0.134	0.046[0.264]0.268	−0.004[0.153]0.154
	PMLE-o	0.001[0.171]0.171	0.000[0.130]0.130	−0.017[0.296]0.296	−0.001[0.134]0.134	0.046[0.264]0.268	−0.004[0.153]0.153
	PMLE-t	0.001[0.171]0.171	0.000[0.130]0.130	−0.017[0.296]0.296	−0.001[0.134]0.134	0.046[0.264]0.268	−0.004[0.153]0.153
200, 2, 0.3	MLE-r	−0.146[0.142]0.204	−0.015[0.145]0.146	−0.055[0.283]0.288	−0.000[0.089]0.089	−2.000[0.000]2.000	−0.300[0.000]0.300
	MLE	0.002[0.187]0.187	−0.014[0.145]0.146	−0.017[0.297]0.297	−0.004[0.132]0.132	0.053[0.273]0.278	−0.007[0.231]0.231
	PMLE-o	0.002[0.187]0.187	−0.014[0.145]0.146	−0.017[0.297]0.297	−0.004[0.133]0.133	0.053[0.273]0.278	−0.007[0.230]0.231
	PMLE-t	0.002[0.187]0.187	−0.014[0.145]0.146	−0.017[0.296]0.297	−0.004[0.133]0.133	0.053[0.273]0.278	−0.006[0.231]0.231
200, 2, −0.3	MLE-r	0.151[0.142]0.208	−0.001[0.143]0.143	−0.054[0.285]0.290	−0.002[0.086]0.086	−2.000[0.000]2.000	0.300[0.000]0.300
	MLE	0.000[0.189]0.189	−0.002[0.144]0.144	−0.016[0.297]0.298	0.002[0.127]0.127	0.050[0.264]0.269	0.003[0.225]0.225
	PMLE-o	0.000[0.189]0.189	−0.002[0.144]0.144	−0.016[0.297]0.298	0.002[0.127]0.127	0.050[0.264]0.269	0.003[0.225]0.225
	PMLE-t	0.000[0.189]0.189	−0.002[0.144]0.144	−0.016[0.297]0.298	0.002[0.127]0.127	0.050[0.264]0.269	0.003[0.225]0.225
200, 2, 0	MLE-r	0.002[0.141]0.141	−0.005[0.140]0.140	−0.051[0.278]0.283	−0.002[0.088]0.088	−2.000[0.000]2.000	0.000[0.000]0.000
	MLE	0.003[0.186]0.186	−0.005[0.142]0.142	−0.036[0.281]0.283	−0.005[0.133]0.133	0.065[0.277]0.285	0.004[0.238]0.238
	PMLE-o	0.003[0.185]0.186	−0.005[0.142]0.142	−0.036[0.281]0.283	−0.006[0.133]0.133	0.065[0.277]0.285	0.003[0.238]0.238
	PMLE-t	0.003[0.185]0.186	−0.005[0.142]0.142	−0.036[0.281]0.283	−0.005[0.133]0.133	0.065[0.277]0.285	0.004[0.238]0.238
200, 0.5, 0.7	MLE-r	−0.174[0.064]0.186	0.003[0.069]0.069	−0.039[0.066]0.076	0.001[0.091]0.091	−2.000[0.000]2.000	−0.700[0.000]0.700
	MLE	0.004[0.082]0.082	0.004[0.066]0.066	−0.003[0.076]0.076	0.001[0.132]0.132	0.066[0.280]0.287	0.012[0.142]0.143
	PMLE-o	0.004[0.082]0.082	0.004[0.066]0.066	−0.003[0.075]0.075	0.001[0.132]0.132	0.067[0.279]0.287	0.012[0.142]0.142
	PMLE-t	0.004[0.082]0.082	0.004[0.066]0.066	−0.003[0.075]0.076	0.001[0.132]0.132	0.067[0.279]0.287	0.012[0.142]0.142
200, 0.5, −0.7	MLE-r	0.177[0.069]0.190	−0.003[0.070]0.070	−0.042[0.070]0.081	0.003[0.090]0.090	−2.000[0.000]2.000	0.700[0.000]0.700
	MLE	0.002[0.082]0.082	−0.004[0.066]0.066	−0.008[0.079]0.080	0.001[0.126]0.126	0.067[0.262]0.270	−0.006[0.137]0.137
	PMLE-o	0.001[0.082]0.082	−0.004[0.066]0.066	−0.008[0.079]0.080	0.001[0.126]0.126	0.067[0.262]0.270	−0.006[0.137]0.137
	PMLE-t	0.002[0.082]0.082	−0.004[0.066]0.066	−0.008[0.079]0.080	0.001[0.126]0.126	0.067[0.262]0.270	−0.006[0.137]0.137
200, 0.5, 0.3	MLE-r	−0.077[0.072]0.105	0.006[0.074]0.074	−0.017[0.068]0.070	−0.000[0.089]0.089	−2.000[0.000]2.000	−0.300[0.000]0.300
	MLE	0.000[0.096]0.096	0.006[0.073]0.074	−0.007[0.071]0.072	0.000[0.132]0.132	0.042[0.264]0.267	0.008[0.220]0.220
	PMLE-o	0.000[0.096]0.096	0.006[0.073]0.074	−0.007[0.071]0.072	0.000[0.132]0.132	0.042[0.264]0.267	0.008[0.220]0.220
	PMLE-t	0.000[0.096]0.096	0.006[0.073]0.074	−0.007[0.071]0.072	0.000[0.132]0.132	0.042[0.264]0.267	0.008[0.220]0.220
200, 0.5, −0.3	MLE-r	0.074[0.073]0.103	−0.002[0.074]0.074	−0.019[0.068]0.071	0.001[0.091]0.091	−2.000[0.000]2.000	0.300[0.000]0.300
	MLE	−0.001[0.094]0.094	−0.001[0.075]0.075	−0.010[0.071]0.072	−0.000[0.130]0.130	0.060[0.281]0.288	0.004[0.224]0.224
	PMLE-o	−0.001[0.094]0.094	−0.002[0.075]0.075	−0.010[0.071]0.072	−0.000[0.130]0.130	0.060[0.281]0.288	0.003[0.223]0.223
	PMLE-t	−0.001[0.094]0.094	−0.002[0.075]0.075	−0.010[0.071]0.072	−0.000[0.130]0.130	0.060[0.281]0.288	0.003[0.223]0.223
200, 0.5, 0	MLE-r	−0.001[0.071]0.071	−0.007[0.075]0.076	−0.011[0.071]0.072	−0.005[0.086]0.086	−2.000[0.000]2.000	0.000[0.000]0.000
	MLE	−0.001[0.092]0.092	−0.007[0.076]0.077	−0.007[0.072]0.073	−0.001[0.135]0.135	0.066[0.279]0.287	−0.001[0.246]0.246
	PMLE-o	−0.001[0.092]0.092	−0.007[0.076]0.077	−0.007[0.072]0.073	−0.001[0.135]0.135	0.066[0.279]0.287	−0.001[0.246]0.246
	PMLE-t	−0.001[0.092]0.092	−0.007[0.076]0.077	−0.007[0.072]0.073	−0.001[0.135]0.135	0.066[0.279]0.287	−0.001[0.246]0.246
600, 2, 0.7	MLE-r	−0.356[0.079]0.364	0.000[0.078]0.078	−0.137[0.159]0.210	−0.000[0.050]0.050	−2.000[0.000]2.000	−0.700[0.000]0.700
	MLE	0.000[0.093]0.093	0.001[0.072]0.072	−0.004[0.180]0.180	0.005[0.069]0.069	0.008[0.147]0.148	0.005[0.080]0.080
	PMLE-o	0.000[0.093]0.093	0.001[0.072]0.072	−0.004[0.180]0.180	0.005[0.069]0.069	0.008[0.147]0.147	0.005[0.080]0.080
	PMLE-t	0.000[0.093]0.093	0.001[0.072]0.072	−0.004[0.180]0.180	0.006[0.069]0.069	0.008[0.147]0.147	0.005[0.080]0.080
600, 2, −0.7	MLE-r	0.351[0.076]0.360	−0.005[0.078]0.078	−0.138[0.154]0.207	0.002[0.051]0.052	−2.000[0.000]2.000	0.700[0.000]0.700
	MLE	−0.001[0.091]0.091	−0.006[0.073]0.073	−0.010[0.175]0.175	0.002[0.073]0.073	0.011[0.148]0.149	−0.002[0.080]0.080
	PMLE-o	−0.001[0.091]0.091	−0.006[0.073]0.073	−0.009[0.175]0.175	0.002[0.073]0.073	0.011[0.148]0.149	−0.002[0.080]0.080
	PMLE-t	−0.001[0.091]0.091	−0.006[0.073]0.073	−0.009[0.175]0.175	0.002[0.073]0.073	0.011[0.148]0.149	−0.002[0.080]0.080
600, 2, 0.3	MLE-r	−0.158[0.081]0.178	−0.000[0.082]0.082	−0.031[0.165]0.168	−0.003[0.052]0.052	−2.000[0.000]2.000	−0.300[0.000]0.300
	MLE	−0.003[0.104]0.104	0.001[0.081]0.081	−0.003[0.170]0.170	−0.001[0.075]0.075	0.019[0.158]0.159	0.006[0.124]0.124
	PMLE-o	−0.003[0.104]0.104	0.001[0.081]0.081	−0.003[0.170]0.170	−0.001[0.075]0.075	0.019[0.158]0.159	0.006[0.124]0.124
	PMLE-t	−0.003[0.104]0.104	0.001[0.081]0.081	−0.003[0.170]0.170	−0.001[0.075]0.075	0.019[0.158]0.159	0.006[0.124]0.124
600, 2, −0.3	MLE-r	0.151[0.084]0.173	−0.000[0.082]0.082	−0.040[0.163]0.168	−0.000[0.051]0.051	−2.000[0.000]2.000	0.300[0.000]0.300
	MLE	−0.001[0.107]0.107	−0.001[0.081]0.081	−0.012[0.168]0.169	−0.002[0.071]0.071	0.018[0.159]0.160	−0.002[0.126]0.126
	PMLE-o	−0.001[0.107]0.107	−0.001[0.081]0.081	−0.012[0.168]0.169	−0.002[0.071]0.071	0.018[0.159]0.160	−0.002[0.126]0.126
	PMLE-t	−0.001[0.107]0.107	−0.001[0.081]0.081	−0.012[0.168]0.169	−0.002[0.071]0.071	0.018[0.159]0.160	−0.002[0.126]0.126
600, 2, 0	MLE-r	−0.005[0.081]0.081	0.002[0.084]0.084	−0.007[0.162]0.162	−0.002[0.050]0.050	−2.000[0.000]2.000	0.000[0.000]0.000
	MLE	−0.005[0.108]0.108	0.002[0.084]0.084	−0.003[0.162]0.162	−0.003[0.074]0.075	0.013[0.151]0.151	0.000[0.131]0.131
	PMLE-o	−0.005[0.108]0.108	0.002[0.084]0.084	−0.003[0.162]0.162	−0.003[0.074]0.075	0.013[0.151]0.151	0.000[0.131]0.131
	PMLE-t	−0.005[0.108]0.108	0.002[0.084]0.084	−0.003[0.162]0.162	−0.003[0.074]0.075	0.013[0.151]0.151	0.000[0.131]0.131
600, 0.5, 0.7	MLE-r	−0.176[0.039]0.180	0.001[0.038]0.038	−0.033[0.039]0.051	−0.000[0.050]0.050	−2.000[0.000]2.000	−0.700[0.000]0.700
	MLE	0.002[0.047]0.047	0.001[0.036]0.036	−0.000[0.043]0.043	−0.000[0.071]0.071	0.014[0.144]0.145	0.005[0.078]0.078
	PMLE-o	0.002[0.047]0.047	0.001[0.036]0.036	−0.000[0.043]0.043	−0.000[0.071]0.071	0.014[0.144]0.145	0.005[0.078]0.078
	PMLE-t	0.002[0.047]0.047	0.001[0.036]0.036	−0.000[0.043]0.043	−0.000[0.071]0.071	0.014[0.144]0.145	0.005[0.078]0.078
600, 0.5, −0.7	MLE-r	0.178[0.040]0.182	−0.000[0.039]0.039	−0.034[0.039]0.052	0.000[0.053]0.053	−2.000[0.000]2.000	0.700[0.000]0.700
	MLE	−0.000[0.048]0.048	−0.000[0.037]0.037	−0.001[0.045]0.045	0.002[0.074]0.074	0.016[0.147]0.148	−0.005[0.080]0.080
	PMLE-o	−0.000[0.048]0.048	−0.000[0.037]0.037	−0.001[0.045]0.045	0.002[0.074]0.075	0.016[0.147]0.148	−0.005[0.080]0.080
	PMLE-t	−0.000[0.048]0.048	−0.000[0.037]0.037	−0.001[0.045]0.045	0.002[0.074]0.075	0.016[0.147]0.148	−0.005[0.080]0.080
600, 0.5, 0.3	MLE-r	−0.075[0.042]0.085	0.002[0.041]0.041	−0.009[0.041]0.042	0.002[0.053]0.053	−2.000[0.000]2.000	−0.300[0.000]0.300
	MLE	0.000[0.053]0.053	0.002[0.041]0.041	−0.002[0.042]0.042	−0.001[0.076]0.076	0.023[0.155]0.156	−0.001[0.124]0.124
	PMLE-o	0.000[0.053]0.053	0.002[0.041]0.041	−0.002[0.042]0.042	−0.001[0.076]0.076	0.023[0.155]0.156	−0.001[0.124]0.124
	PMLE-t	0.000[0.053]0.053	0.002[0.041]0.041	−0.002[0.042]0.042	−0.001[0.076]0.076	0.023[0.155]0.156	−0.001[0.124]0.124
600, 0.5, −0.3	MLE-r	0.076[0.039]0.085	0.000[0.041]0.041	−0.012[0.041]0.043	−0.002[0.052]0.052	−2.000[0.000]2.000	0.300[0.000]0.300
	MLE	0.001[0.051]0.051	0.000[0.040]0.040	−0.005[0.043]0.043	−0.005[0.074]0.075	0.019[0.156]0.157	0.002[0.121]0.121
	PMLE-o	0.001[0.051]0.051	0.000[0.040]0.040	−0.005[0.043]0.043	−0.005[0.074]0.075	0.019[0.156]0.157	0.002[0.121]0.121
	PMLE-t	0.001[0.051]0.051	0.000[0.040]0.040	−0.005[0.043]0.043	−0.005[0.074]0.075	0.019[0.156]0.157	0.002[0.121]0.121
600, 0.5, 0	MLE-r	−0.001[0.040]0.040	0.001[0.041]0.041	−0.003[0.041]0.041	−0.002[0.052]0.052	−2.000[0.000]2.000	0.000[0.000]0.000
	MLE	−0.000[0.052]0.052	0.001[0.041]0.041	−0.002[0.041]0.041	−0.005[0.074]0.074	0.015[0.146]0.147	0.001[0.129]0.129
	PMLE-o	−0.000[0.052]0.052	0.001[0.041]0.041	−0.002[0.041]0.041	−0.005[0.074]0.074	0.015[0.146]0.147	0.001[0.129]0.129
	PMLE-t	−0.000[0.052]0.052	0.001[0.041]0.041	−0.002[0.041]0.041	−0.005[0.074]0.074	0.015[0.146]0.147	0.001[0.129]0.129

The maximum likelihood estimator (MLE)-r denotes the restricted MLE with the restriction $\theta_2 = 0$ imposed, and the PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias [SE]RMSE. $(\beta_{10}, \beta_{20}, \gamma_{10}) = (1, 1, 1)$.

Table 3. The biases, SEs and RMSEs of the estimators when $\gamma_{20} = 0.5$ in the sample selection model.

n, σ_0^2, ρ_0		β_1	β_2	σ^2	γ_1	γ_2	ρ
200, 2, 0.7	MLE-r	−0.704[0.123]0.715	0.001[0.125]0.125	−0.532[0.204]0.570	0.002[0.090]0.090	−0.500[0.000]0.500	−0.700[0.000]0.700
	MLE	−0.014[0.323]0.323	0.003[0.120]0.120	0.027[0.483]0.484	0.003[0.094]0.094	0.006[0.098]0.098	−0.035[0.217]0.220
	PMLE-o	−0.015[0.324]0.325	0.003[0.120]0.120	0.028[0.487]0.487	0.003[0.094]0.094	0.006[0.099]0.099	−0.035[0.218]0.221
	PMLE-t	−0.015[0.324]0.324	0.003[0.120]0.120	0.027[0.483]0.484	0.003[0.094]0.094	0.006[0.099]0.099	−0.035[0.218]0.221
200, 2, −0.7	MLE-r	0.705[0.124]0.716	0.003[0.125]0.125	−0.533[0.208]0.572	−0.004[0.093]0.093	−0.500[0.000]0.500	0.700[0.000]0.700
	MLE	0.009[0.306]0.306	0.001[0.123]0.123	0.030[0.508]0.509	−0.003[0.097]0.097	0.012[0.104]0.104	0.031[0.207]0.209
	PMLE-o	0.009[0.308]0.308	0.001[0.123]0.123	0.031[0.510]0.511	−0.002[0.097]0.097	0.012[0.104]0.104	0.031[0.207]0.209
	PMLE-t	0.009[0.306]0.306	0.001[0.123]0.123	0.030[0.508]0.509	−0.003[0.097]0.097	0.012[0.104]0.104	0.031[0.207]0.209
200, 2, 0.3	MLE-r	−0.305[0.140]0.336	0.003[0.142]0.142	−0.126[0.270]0.297	0.000[0.088]0.088	−0.500[0.000]0.500	−0.300[0.000]0.300
	MLE	0.014[0.404]0.404	0.003[0.142]0.142	0.124[0.486]0.501	0.002[0.092]0.092	0.006[0.102]0.102	−0.009[0.324]0.324
	PMLE-o	0.017[0.410]0.410	0.002[0.142]0.142	0.130[0.506]0.523	0.002[0.092]0.092	0.006[0.103]0.103	−0.007[0.325]0.325
	PMLE-t	0.014[0.404]0.404	0.003[0.142]0.142	0.123[0.486]0.501	0.002[0.092]0.092	0.006[0.103]0.103	−0.009[0.324]0.324
200, 2, −0.3	MLE-r	0.301[0.139]0.331	0.002[0.142]0.142	−0.124[0.273]0.300	0.002[0.089]0.089	−0.500[0.000]0.500	0.300[0.000]0.300
	MLE	−0.014[0.421]0.421	0.002[0.142]0.142	0.125[0.472]0.489	0.002[0.094]0.094	0.010[0.107]0.107	0.012[0.332]0.333
	PMLE-o	−0.014[0.420]0.421	0.002[0.142]0.142	0.125[0.472]0.488	0.002[0.094]0.094	0.009[0.109]0.110	0.012[0.332]0.332
	PMLE-t	−0.015[0.421]0.421	0.002[0.142]0.142	0.125[0.472]0.489	0.002[0.094]0.094	0.009[0.107]0.108	0.012[0.332]0.332
200, 2, 0	MLE-r	0.004[0.144]0.144	0.005[0.147]0.147	−0.051[0.275]0.280	−0.004[0.092]0.093	−0.500[0.000]0.500	0.000[0.000]0.000
	MLE	0.007[0.446]0.446	0.006[0.148]0.149	0.124[0.401]0.419	−0.002[0.098]0.098	0.010[0.105]0.105	0.001[0.370]0.370
	PMLE-o	0.007[0.446]0.446	0.006[0.148]0.149	0.123[0.400]0.419	−0.002[0.098]0.098	0.010[0.106]0.106	0.002[0.369]0.369
	PMLE-t	0.007[0.446]0.446	0.006[0.148]0.149	0.123[0.400]0.419	−0.002[0.098]0.098	0.010[0.106]0.106	0.002[0.369]0.369
200, 0.5, 0.7	MLE-r	−0.356[0.059]0.361	0.002[0.064]0.064	−0.130[0.054]0.141	−0.000[0.090]0.090	−0.500[0.000]0.500	−0.700[0.000]0.700
	MLE	−0.009[0.158]0.158	0.002[0.065]0.065	0.012[0.127]0.127	−0.001[0.097]0.097	0.016[0.103]0.104	−0.033[0.227]0.229
	PMLE-o	−0.009[0.158]0.158	0.002[0.065]0.065	0.012[0.127]0.127	−0.001[0.097]0.097	0.016[0.103]0.104	−0.033[0.227]0.229
	PMLE-t	−0.009[0.158]0.158	0.002[0.065]0.065	0.012[0.127]0.127	−0.001[0.097]0.097	0.016[0.103]0.104	−0.033[0.227]0.229
200, 0.5, −0.7	MLE-r	0.350[0.063]0.356	−0.000[0.062]0.062	−0.133[0.056]0.144	−0.003[0.088]0.088	−0.500[0.000]0.500	0.700[0.000]0.700
	MLE	0.009[0.147]0.147	−0.000[0.060]0.060	0.001[0.125]0.125	−0.001[0.093]0.093	0.017[0.102]0.104	0.038[0.204]0.207
	PMLE-o	0.010[0.151]0.151	−0.000[0.061]0.061	0.002[0.125]0.126	−0.001[0.093]0.093	0.017[0.103]0.104	0.039[0.210]0.214
	PMLE-t	0.009[0.147]0.147	−0.000[0.060]0.060	0.001[0.125]0.125	−0.001[0.093]0.093	0.017[0.102]0.104	0.038[0.204]0.207
200, 0.5, 0.3	MLE-r	−0.145[0.070]0.161	−0.000[0.068]0.068	−0.035[0.068]0.076	0.005[0.090]0.090	−0.500[0.000]0.500	−0.300[0.000]0.300
	MLE	0.006[0.212]0.212	−0.000[0.069]0.069	0.027[0.123]0.126	0.003[0.096]0.096	0.007[0.109]0.110	−0.028[0.338]0.339
	PMLE-o	0.006[0.212]0.212	−0.000[0.069]0.069	0.028[0.123]0.126	0.003[0.096]0.096	0.006[0.111]0.111	−0.028[0.338]0.339
	PMLE-t	0.006[0.212]0.212	−0.000[0.069]0.069	0.028[0.123]0.126	0.003[0.096]0.096	0.007[0.109]0.110	−0.028[0.338]0.339
200, 0.5, −0.3	MLE-r	0.152[0.068]0.167	0.003[0.070]0.070	−0.032[0.065]0.072	0.004[0.088]0.088	−0.500[0.000]0.500	0.300[0.000]0.300
	MLE	0.009[0.203]0.203	0.003[0.071]0.071	0.025[0.105]0.108	0.003[0.092]0.092	0.010[0.106]0.106	0.036[0.331]0.333
	PMLE-o	0.010[0.202]0.202	0.003[0.071]0.071	0.024[0.105]0.108	0.003[0.092]0.092	0.009[0.108]0.108	0.036[0.330]0.332
	PMLE-t	0.010[0.202]0.202	0.003[0.071]0.071	0.024[0.105]0.108	0.003[0.092]0.092	0.009[0.108]0.108	0.036[0.330]0.332
200, 0.5, 0	MLE-r	−0.000[0.072]0.072	−0.001[0.070]0.070	−0.010[0.071]0.072	−0.001[0.086]0.086	−0.500[0.000]0.500	0.000[0.000]0.000
	MLE	0.004[0.216]0.216	−0.002[0.071]0.071	0.032[0.104]0.108	−0.001[0.090]0.090	0.005[0.107]0.107	0.007[0.360]0.360
	PMLE-o	0.003[0.219]0.219	−0.002[0.071]0.071	0.033[0.107]0.112	−0.001[0.090]0.090	0.004[0.110]0.110	0.006[0.363]0.363
	PMLE-t	0.005[0.215]0.216	−0.002[0.071]0.071	0.031[0.104]0.108	−0.001[0.090]0.090	0.005[0.108]0.109	0.008[0.360]0.360
600, 2, 0.7	MLE-r	−0.707[0.072]0.711	−0.004[0.069]0.069	−0.506[0.124]0.521	0.001[0.050]0.050	−0.500[0.000]0.500	−0.700[0.000]0.700
	MLE	−0.003[0.172]0.172	−0.004[0.066]0.066	0.015[0.287]0.288	0.001[0.052]0.052	−0.000[0.059]0.059	−0.010[0.106]0.106
	PMLE-o	−0.003[0.172]0.172	−0.004[0.066]0.066	0.015[0.287]0.288	0.001[0.052]0.052	−0.000[0.059]0.059	−0.010[0.106]0.106
	PMLE-t	−0.003[0.172]0.172	−0.004[0.066]0.066	0.015[0.287]0.288	0.001[0.052]0.052	−0.000[0.059]0.059	−0.010[0.106]0.106
600, 2, −0.7	MLE-r	0.709[0.071]0.712	0.002[0.070]0.070	−0.518[0.124]0.533	−0.002[0.050]0.050	−0.500[0.000]0.500	0.700[0.000]0.700
	MLE	0.010[0.166]0.166	0.002[0.069]0.069	−0.009[0.279]0.280	−0.002[0.052]0.052	0.005[0.056]0.057	0.012[0.106]0.106
	PMLE-o	0.010[0.166]0.166	0.002[0.069]0.069	−0.009[0.279]0.280	−0.002[0.052]0.052	0.005[0.056]0.057	0.012[0.106]0.106
	PMLE-t	0.010[0.166]0.166	0.002[0.069]0.069	−0.009[0.279]0.280	−0.002[0.052]0.052	0.005[0.056]0.057	0.012[0.106]0.106
600, 2, 0.3	MLE-r	−0.303[0.079]0.313	0.000[0.081]0.081	−0.100[0.163]0.191	−0.001[0.050]0.050	−0.500[0.000]0.500	−0.300[0.000]0.300
	MLE	0.009[0.231]0.231	0.001[0.081]0.081	0.044[0.239]0.243	0.000[0.052]0.052	0.003[0.058]0.058	−0.001[0.196]0.196
	PMLE-o	0.009[0.231]0.231	0.001[0.081]0.081	0.044[0.239]0.243	0.000[0.052]0.052	0.003[0.058]0.058	−0.001[0.196]0.196
	PMLE-t	0.009[0.231]0.231	0.001[0.081]0.081	0.044[0.239]0.243	0.000[0.052]0.052	0.003[0.058]0.058	−0.001[0.196]0.196
600, 2, −0.3	MLE-r	0.305[0.082]0.316	−0.001[0.079]0.079	−0.104[0.158]0.189	−0.000[0.053]0.053	−0.500[0.000]0.500	0.300[0.000]0.300
	MLE	0.012[0.229]0.229	−0.000[0.079]0.079	0.028[0.228]0.229	−0.000[0.055]0.055	0.002[0.057]0.057	0.018[0.196]0.197
	PMLE-o	0.012[0.229]0.229	−0.000[0.079]0.079	0.028[0.228]0.229	−0.000[0.055]0.055	0.002[0.057]0.057	0.018[0.196]0.197
	PMLE-t	0.012[0.229]0.229	−0.000[0.079]0.079	0.028[0.228]0.229	−0.000[0.055]0.055	0.002[0.057]0.057	0.018[0.196]0.197
600, 2, 0	MLE-r	0.002[0.082]0.082	0.001[0.082]0.082	−0.011[0.154]0.155	0.004[0.050]0.051	−0.500[0.000]0.500	0.000[0.000]0.000
	MLE	−0.003[0.226]0.226	0.001[0.082]0.082	0.035[0.176]0.180	0.003[0.051]0.052	0.001[0.061]0.061	−0.005[0.205]0.205
	PMLE-o	−0.003[0.226]0.226	0.001[0.082]0.082	0.035[0.176]0.180	0.003[0.051]0.052	0.001[0.061]0.061	−0.005[0.205]0.205
	PMLE-t	−0.003[0.226]0.226	0.001[0.082]0.082	0.035[0.176]0.180	0.003[0.051]0.052	0.001[0.061]0.061	−0.005[0.205]0.205
600, 0.5, 0.7	MLE-r	−0.351[0.036]0.353	0.000[0.035]0.035	−0.127[0.031]0.130	−0.001[0.050]0.050	−0.500[0.000]0.500	−0.700[0.000]0.700
	MLE	−0.001[0.084]0.084	−0.001[0.034]0.034	0.002[0.070]0.070	−0.001[0.053]0.053	0.002[0.059]0.059	−0.013[0.109]0.110
	PMLE-o	−0.001[0.084]0.084	−0.001[0.034]0.034	0.002[0.070]0.070	−0.001[0.053]0.053	0.002[0.059]0.059	−0.013[0.109]0.110
	PMLE-t	−0.001[0.084]0.084	−0.001[0.034]0.034	0.002[0.070]0.070	−0.001[0.053]0.053	0.002[0.059]0.059	−0.013[0.109]0.110
600, 0.5, −0.7	MLE-r	0.353[0.036]0.355	0.001[0.036]0.036	−0.128[0.030]0.131	−0.002[0.051]0.051	−0.500[0.000]0.500	0.700[0.000]0.700
	MLE	0.002[0.081]0.081	0.001[0.034]0.034	0.001[0.069]0.069	−0.002[0.054]0.054	0.003[0.058]0.058	0.010[0.101]0.101
	PMLE-o	0.002[0.081]0.081	0.001[0.034]0.034	0.001[0.069]0.069	−0.002[0.054]0.054	0.003[0.058]0.058	0.010[0.101]0.101
	PMLE-t	0.002[0.081]0.081	0.001[0.034]0.034	0.001[0.069]0.069	−0.002[0.054]0.054	0.003[0.058]0.058	0.010[0.101]0.101
600, 0.5, 0.3	MLE-r	−0.149[0.040]0.154	−0.002[0.039]0.039	−0.025[0.039]0.047	0.002[0.051]0.051	−0.500[0.000]0.500	−0.300[0.000]0.300
	MLE	0.006[0.114]0.114	−0.002[0.039]0.039	0.010[0.057]0.058	0.002[0.054]0.054	0.003[0.059]0.059	−0.000[0.188]0.188
	PMLE-o	0.006[0.114]0.114	−0.002[0.039]0.039	0.010[0.057]0.058	0.002[0.054]0.054	0.003[0.059]0.059	−0.000[0.188]0.188
	PMLE-t	0.006[0.114]0.114	−0.002[0.039]0.039	0.010[0.057]0.058	0.002[0.054]0.054	0.003[0.059]0.059	−0.000[0.188]0.188
600, 0.5, −0.3	MLE-r	0.152[0.039]0.157	−0.002[0.040]0.040	−0.026[0.040]0.047	0.001[0.053]0.053	−0.500[0.000]0.500	0.300[0.000]0.300
	MLE	0.003[0.110]0.110	−0.002[0.040]0.040	0.006[0.056]0.057	0.000[0.055]0.055	0.002[0.059]0.060	0.014[0.188]0.188
	PMLE-o	0.003[0.110]0.110	−0.002[0.040]0.040	0.006[0.056]0.057	0.000[0.055]0.055	0.002[0.059]0.060	0.014[0.188]0.188
	PMLE-t	0.003[0.110]0.110	−0.002[0.040]0.040	0.006[0.056]0.057	0.000[0.055]0.055	0.002[0.059]0.060	0.014[0.188]0.188
600, 0.5, 0	MLE-r	0.001[0.040]0.040	0.000[0.042]0.042	−0.004[0.041]0.042	−0.001[0.052]0.052	−0.500[0.000]0.500	0.000[0.000]0.000
	MLE	−0.003[0.119]0.119	0.000[0.042]0.042	0.008[0.047]0.047	−0.001[0.053]0.053	0.002[0.060]0.060	−0.007[0.212]0.213
	PMLE-o	−0.003[0.119]0.119	0.000[0.042]0.042	0.008[0.047]0.047	−0.001[0.053]0.053	0.002[0.060]0.060	−0.007[0.212]0.213
	PMLE-t	−0.003[0.119]0.119	0.000[0.042]0.042	0.008[0.047]0.047	−0.001[0.053]0.053	0.002[0.060]0.060	−0.007[0.212]0.213

The MLE-r denotes the restricted MLE with the restriction $\theta_2 = 0$ imposed, and the PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias[SE]RMSE. $(\beta_{10}, \beta_{20}, \gamma_{10}) = (1, 1, 1)$.

Table 4. The biases, SEs and RMSEs of the estimators when $\gamma_{20} = 0$ and $\rho_0 \neq 0$ in the sample selection model.

n, σ_0^2, ρ_0		β_1	β_2	σ^2	γ_1	γ_2	ρ
200, 2, 0.7	MLE-r	−0.792[0.118]0.801	−0.002[0.121]0.121	−0.647[0.199]0.677	−0.001[0.091]0.091	0.000[0.000]0.000	−0.700[0.000]0.700
	MLE	−0.482[0.884]1.007	−0.001[0.124]0.124	0.210[0.666]0.698	−0.002[0.092]0.092	−0.004[0.096]0.096	−0.459[0.698]0.835
	PMLE-o	−0.743[0.333]0.814	−0.001[0.122]0.122	−0.546[0.510]0.747	−0.002[0.099]0.099	−0.000[0.036]0.036	−0.666[0.215]0.699
	PMLE-t	−0.665[0.521]0.845	−0.001[0.122]0.122	−0.376[0.639]0.741	−0.001[0.091]0.091	−0.001[0.050]0.050	−0.606[0.382]0.717
200, 2, −0.7	MLE-r	0.786[0.114]0.794	0.004[0.115]0.115	−0.649[0.191]0.676	−0.001[0.089]0.089	0.000[0.000]0.000	0.700[0.000]0.700
	MLE	0.420[0.867]0.963	0.004[0.116]0.117	0.213[0.650]0.684	−0.001[0.090]0.090	−0.001[0.098]0.098	0.421[0.687]0.806
	PMLE-o	0.735[0.326]0.804	0.004[0.115]0.115	−0.550[0.462]0.718	−0.000[0.090]0.090	−0.002[0.043]0.043	0.664[0.226]0.701
	PMLE-t	0.648[0.538]0.842	0.004[0.116]0.116	−0.361[0.636]0.731	−0.001[0.089]0.089	−0.003[0.055]0.055	0.598[0.396]0.718
200, 2, 0.3	MLE-r	−0.343[0.136]0.369	0.008[0.139]0.139	−0.158[0.263]0.307	0.000[0.093]0.093	0.000[0.000]0.000	−0.300[0.000]0.300
	MLE	−0.302[1.047]1.089	0.008[0.143]0.143	0.908[0.847]1.242	−0.000[0.093]0.093	0.005[0.098]0.099	−0.270[0.719]0.768
	PMLE-o	−0.340[0.332]0.475	0.009[0.141]0.141	−0.071[0.555]0.559	0.000[0.099]0.099	−0.001[0.037]0.037	−0.296[0.180]0.347
	PMLE-t	−0.326[0.569]0.656	0.008[0.141]0.141	0.145[0.789]0.802	0.000[0.093]0.093	−0.001[0.051]0.051	−0.289[0.362]0.463
200, 2, −0.3	MLE-r	0.340[0.142]0.368	0.001[0.138]0.138	−0.161[0.261]0.307	0.002[0.089]0.089	0.000[0.000]0.000	0.300[0.000]0.300
	MLE	0.347[1.029]1.086	0.001[0.142]0.142	0.878[0.827]1.206	0.002[0.091]0.091	0.001[0.102]0.102	0.304[0.712]0.774
	PMLE-o	0.353[0.285]0.454	0.001[0.139]0.139	−0.095[0.466]0.476	0.001[0.094]0.094	0.001[0.041]0.041	0.309[0.164]0.350
	PMLE-t	0.376[0.567]0.680	0.001[0.139]0.139	0.142[0.777]0.790	0.002[0.090]0.090	−0.000[0.054]0.054	0.323[0.361]0.484
200, 0.5, 0.7	MLE-r	−0.397[0.061]0.402	−0.001[0.060]0.060	−0.161[0.048]0.168	0.001[0.091]0.091	0.000[0.000]0.000	−0.700[0.000]0.700
	MLE	−0.240[0.425]0.488	−0.001[0.061]0.061	0.037[0.158]0.163	0.001[0.091]0.091	0.002[0.100]0.100	−0.464[0.679]0.822
	PMLE-o	−0.378[0.152]0.408	−0.001[0.060]0.060	−0.142[0.111]0.180	0.001[0.094]0.094	0.002[0.040]0.040	−0.673[0.192]0.700
	PMLE-t	−0.341[0.242]0.418	−0.001[0.060]0.060	−0.105[0.148]0.181	0.001[0.091]0.091	0.002[0.054]0.054	−0.617[0.347]0.708
200, 0.5, -0.7	MLE-r	0.397[0.059]0.401	0.004[0.060]0.060	−0.166[0.048]0.173	−0.001[0.093]0.093	0.000[0.000]0.000	0.700[0.000]0.700
	MLE	0.241[0.434]0.496	0.004[0.060]0.061	0.039[0.161]0.166	−0.001[0.093]0.093	0.001[0.097]0.097	0.463[0.692]0.833
	PMLE-o	0.382[0.139]0.407	0.004[0.060]0.060	−0.151[0.096]0.179	−0.001[0.094]0.094	0.001[0.040]0.040	0.679[0.180]0.702
	PMLE-t	0.342[0.247]0.422	0.004[0.060]0.060	−0.107[0.148]0.182	−0.000[0.093]0.093	0.002[0.054]0.054	0.618[0.364]0.717
200, 0.5, 0.3	MLE-r	−0.171[0.070]0.184	0.002[0.071]0.071	−0.042[0.066]0.078	0.004[0.093]0.093	0.000[0.000]0.000	−0.300[0.000]0.300
	MLE	−0.173[0.518]0.547	0.002[0.072]0.072	0.220[0.209]0.304	0.004[0.094]0.094	−0.003[0.104]0.104	−0.308[0.715]0.779
	PMLE-o	−0.167[0.174]0.241	0.002[0.071]0.071	−0.017[0.139]0.140	0.003[0.095]0.095	−0.001[0.043]0.043	−0.294[0.198]0.354
	PMLE-t	−0.158[0.288]0.328	0.002[0.071]0.071	0.037[0.192]0.196	0.004[0.093]0.093	−0.002[0.055]0.056	−0.285[0.374]0.470
200, 0.5, −0.3	MLE-r	0.167[0.071]0.181	0.002[0.069]0.069	−0.039[0.068]0.079	−0.003[0.088]0.088	0.000[0.000]0.000	0.300[0.000]0.300
	MLE	0.157[0.517]0.540	0.002[0.070]0.070	0.222[0.216]0.309	−0.003[0.089]0.089	0.002[0.100]0.100	0.285[0.711]0.766
	PMLE-o	0.164[0.154]0.225	0.002[0.069]0.069	−0.020[0.144]0.145	−0.002[0.090]0.090	0.000[0.034]0.034	0.295[0.159]0.336
	PMLE-t	0.174[0.269]0.320	0.002[0.069]0.069	0.028[0.196]0.198	−0.003[0.088]0.088	−0.000[0.046]0.046	0.306[0.335]0.454
600, 2, 0.7	MLE-r	−0.786[0.068]0.789	−0.002[0.070]0.070	−0.634[0.114]0.645	0.001[0.051]0.051	0.000[0.000]0.000	−0.700[0.000]0.700
	MLE	−0.322[0.645]0.721	−0.002[0.070]0.070	0.001[0.415]0.415	0.001[0.051]0.051	0.002[0.053]0.053	−0.317[0.561]0.644
	PMLE-o	−0.771[0.147]0.785	−0.001[0.070]0.070	−0.616[0.202]0.648	0.001[0.061]0.061	0.000[0.011]0.011	−0.688[0.100]0.695
	PMLE-t	−0.581[0.466]0.745	−0.002[0.070]0.070	−0.378[0.452]0.589	0.001[0.051]0.051	0.001[0.029]0.029	−0.531[0.384]0.656
600, 2, −0.7	MLE-r	0.788[0.068]0.790	0.002[0.069]0.069	−0.637[0.114]0.647	−0.001[0.049]0.049	0.000[0.000]0.000	0.700[0.000]0.700
	MLE	0.281[0.623]0.683	0.002[0.069]0.069	0.007[0.408]0.408	−0.001[0.050]0.050	−0.000[0.050]0.050	0.280[0.539]0.607
	PMLE-o	0.779[0.116]0.787	0.001[0.069]0.069	−0.627[0.164]0.648	0.000[0.048]0.048	0.000[0.005]0.005	0.694[0.075]0.698
	PMLE-t	0.572[0.465]0.737	0.002[0.069]0.069	−0.378[0.435]0.576	−0.001[0.049]0.049	−0.000[0.029]0.029	0.522[0.388]0.651
600, 2, 0.3	MLE-r	−0.339[0.079]0.348	0.001[0.082]0.082	−0.132[0.152]0.201	0.001[0.052]0.052	0.000[0.000]0.000	−0.300[0.000]0.300
	MLE	−0.326[0.844]0.904	0.001[0.083]0.083	0.565[0.508]0.760	0.001[0.052]0.052	0.002[0.057]0.057	−0.290[0.626]0.690
	PMLE-o	−0.341[0.106]0.357	0.002[0.082]0.082	−0.127[0.174]0.216	0.001[0.060]0.060	0.000[0.009]0.009	−0.301[0.046]0.304
	PMLE-t	−0.330[0.504]0.602	0.001[0.083]0.083	0.113[0.516]0.528	0.001[0.052]0.052	0.002[0.029]0.029	−0.294[0.358]0.463
600, 2, −0.3	MLE-r	0.343[0.075]0.351	−0.002[0.082]0.082	−0.122[0.151]0.194	0.001[0.052]0.052	0.000[0.000]0.000	0.300[0.000]0.300
	MLE	0.311[0.838]0.894	−0.002[0.082]0.082	0.564[0.509]0.760	0.001[0.052]0.052	−0.002[0.055]0.055	0.279[0.621]0.681
	PMLE-o	0.342[0.102]0.357	−0.002[0.081]0.081	−0.118[0.183]0.217	0.001[0.054]0.054	−0.000[0.007]0.007	0.300[0.042]0.303
	PMLE-t	0.335[0.492]0.595	−0.002[0.081]0.081	0.111[0.499]0.511	0.001[0.052]0.052	−0.001[0.032]0.032	0.296[0.352]0.460
600, 0.5, 0.7	MLE-r	−0.394[0.035]0.396	−0.000[0.036]0.036	−0.159[0.028]0.161	−0.001[0.051]0.051	0.000[0.000]0.000	−0.700[0.000]0.700
	MLE	−0.159[0.323]0.360	−0.000[0.036]0.036	−0.001[0.107]0.107	−0.001[0.051]0.051	0.002[0.054]0.054	−0.312[0.552]0.634
	PMLE-o	−0.390[0.064]0.395	−0.000[0.036]0.036	−0.156[0.044]0.162	−0.001[0.056]0.056	0.000[0.009]0.009	−0.693[0.076]0.697
	PMLE-t	−0.304[0.222]0.377	−0.000[0.036]0.036	−0.102[0.111]0.151	−0.001[0.051]0.051	−0.001[0.028]0.028	−0.554[0.362]0.662
600, 0.5, −0.7	MLE-r	0.394[0.033]0.396	−0.000[0.034]0.034	−0.158[0.028]0.160	−0.002[0.050]0.050	0.000[0.000]0.000	0.700[0.000]0.700
	MLE	0.148[0.313]0.347	−0.000[0.034]0.034	−0.002[0.106]0.106	−0.002[0.050]0.050	0.002[0.055]0.055	0.293[0.535]0.610
	PMLE-o	0.389[0.067]0.394	−0.001[0.034]0.034	−0.154[0.046]0.161	−0.003[0.050]0.050	−0.000[0.011]0.011	0.691[0.088]0.697
	PMLE-t	0.302[0.210]0.368	−0.001[0.034]0.034	−0.107[0.106]0.151	−0.002[0.050]0.050	0.000[0.027]0.027	0.549[0.339]0.645
600, 0.5, 0.3	MLE-r	−0.169[0.041]0.174	−0.003[0.040]0.040	−0.030[0.039]0.050	−0.000[0.050]0.050	0.000[0.000]0.000	−0.300[0.000]0.300
	MLE	−0.154[0.413]0.441	−0.004[0.040]0.040	0.138[0.125]0.186	−0.000[0.051]0.051	−0.000[0.055]0.055	−0.281[0.616]0.677
	PMLE-o	−0.169[0.048]0.176	−0.003[0.040]0.040	−0.030[0.042]0.052	−0.001[0.053]0.053	0.000[0.006]0.006	−0.300[0.034]0.302
	PMLE-t	−0.163[0.235]0.285	−0.004[0.040]0.040	0.023[0.120]0.122	−0.000[0.050]0.050	0.000[0.027]0.027	−0.293[0.336]0.446
600, 0.5, −0.3	MLE-r	0.170[0.039]0.174	−0.000[0.040]0.040	−0.031[0.039]0.050	−0.001[0.051]0.051	0.000[0.000]0.000	0.300[0.000]0.300
	MLE	0.148[0.422]0.447	−0.001[0.040]0.040	0.145[0.127]0.193	−0.001[0.051]0.051	0.000[0.053]0.053	0.268[0.627]0.682
	PMLE-o	0.170[0.044]0.176	−0.000[0.040]0.040	−0.030[0.041]0.051	−0.002[0.052]0.052	0.000[0.002]0.002	0.301[0.028]0.302
	PMLE-t	0.166[0.225]0.280	−0.000[0.040]0.040	0.018[0.116]0.118	−0.001[0.051]0.051	0.000[0.023]0.023	0.295[0.323]0.438

The MLE-r denotes the restricted MLE with the restriction $\theta_2 = 0$ imposed, and the PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias[SE]RMSE. $(\beta_{10}, \beta_{20}, \gamma_{10}) = (1, 1, 1)$.

Table 5. The biases, SEs and RMSEs of the estimators when $\gamma_{20} = 0$ and $\rho_0 = 0$ in the sample selection model.

n, σ_0^2		β_1	β_2	σ^2	γ_1	γ_2	ρ
200, 2	MLE-r	0.003[0.141]0.141	−0.005[0.136]0.137	−0.040[0.284]0.287	0.003[0.092]0.092	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	0.000[1.076]1.076	−0.004[0.138]0.138	1.062[0.887]1.383	0.002[0.093]0.093	−0.001[0.100]0.100	−0.004[0.712]0.712
	PMLE-o	0.001[0.319]0.319	−0.004[0.137]0.137	0.041[0.521]0.522	0.001[0.098]0.098	−0.001[0.041]0.041	0.000[0.176]0.176
	PMLE-t	0.024[0.581]0.581	−0.005[0.137]0.137	0.271[0.801]0.845	0.003[0.092]0.092	−0.001[0.053]0.053	0.014[0.359]0.359
200, 0.5	MLE-r	0.004[0.071]0.072	0.000[0.073]0.073	−0.014[0.074]0.075	−0.002[0.086]0.086	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	0.012[0.535]0.535	−0.001[0.074]0.074	0.261[0.232]0.349	−0.002[0.087]0.087	−0.002[0.101]0.101	0.012[0.709]0.709
	PMLE-o	0.001[0.156]0.156	0.000[0.074]0.074	0.005[0.135]0.135	−0.003[0.089]0.089	−0.001[0.031]0.031	−0.004[0.164]0.164
	PMLE-t	0.009[0.290]0.290	0.000[0.074]0.074	0.061[0.200]0.209	−0.002[0.086]0.086	−0.002[0.048]0.048	0.006[0.353]0.353
600, 2	MLE-r	0.002[0.082]0.082	−0.006[0.081]0.081	−0.018[0.167]0.168	−0.001[0.051]0.051	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	−0.014[0.864]0.864	−0.006[0.082]0.082	0.713[0.537]0.893	−0.001[0.051]0.051	−0.001[0.056]0.056	−0.011[0.623]0.623
	PMLE-o	0.002[0.115]0.115	−0.006[0.081]0.081	−0.011[0.211]0.211	−0.001[0.057]0.057	−0.000[0.008]0.008	0.000[0.049]0.049
	PMLE-t	0.017[0.539]0.539	−0.006[0.081]0.081	0.261[0.547]0.606	−0.001[0.051]0.051	0.000[0.032]0.032	0.010[0.375]0.375
600, 0.5	MLE-r	0.001[0.041]0.041	0.002[0.041]0.041	−0.003[0.040]0.040	−0.001[0.051]0.051	0.000[0.000]0.000	0.000[0.000]0.000
	MLE	0.025[0.437]0.438	0.001[0.041]0.041	0.185[0.134]0.229	−0.001[0.051]0.051	−0.002[0.056]0.056	0.033[0.629]0.630
	PMLE-o	−0.000[0.053]0.053	0.002[0.041]0.041	−0.002[0.046]0.046	−0.001[0.053]0.053	0.000[0.007]0.007	−0.001[0.046]0.046
	PMLE-t	0.013[0.250]0.250	0.001[0.041]0.041	0.057[0.131]0.143	−0.001[0.051]0.051	0.000[0.028]0.028	0.015[0.346]0.346

The MLE-r denotes the restricted MLE with the restriction $\theta_2 = 0$ imposed, and the PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias[SE]RMSE. $(\beta_{10}, \beta_{20}, \gamma_{10}) = (1, 1, 1)$.

4.2. The Stochastic Frontier Function Model

In the Monte Carlo experiments for the stochastic frontier function model, there are three explanatory variables in x: the first one is the intercept term, the second one is randomly drawn from the standard normal distribution, and the third one is randomly drawn from the centered chi-squared distribution $\chi^2(2) - 2$. The true coefficient vector β_0 for the explanatory variables is $(1, 1, 1)'$. We fix $\sigma_{20}^2 = 1$, thus $\sigma_0^2 = \delta_0^2 + 1$, where δ_0 is either 2, 1, 0.5, 0.25, 0.1 or 0. For the PML criterion function (1) using the original likelihood function, μ is set to 4, and Γ_n in the information criterion (2) is taken to be $\Gamma_n = 0.1n^{-1/2}$, which is chosen in a way similar to that for the sample selection model. For the PML criterion function using the transformed likelihood function as in (5), $3\mu_1 = 4$ and $\Gamma_n = 0.1n^{-1/2}$.

Table 6 reports the probabilities that the PMLEs select the right model. For sample size $n = 200$, when $\delta_0 = 2$, both the PMLE-o and PMLE-t estimate δ to be nonzero with probabilities higher than 80%. However, when $\delta_0 = 1$, 0.5, 0.25 or 0.1, the PMLEs estimate δ to be nonzero with very low probabilities. With $\delta_0 = 0$, the PMLEs estimate δ as zero with probabilities higher than 85%. There is a weak identification issue for the stochastic frontier function model similar to that for the sample selection model: $\psi_0\sigma_0\frac{\partial L_n(\theta_{10},\delta)}{\partial\beta_1} + 2\sigma_0^2\psi_0^2\delta\frac{\partial L_n(\theta_{10},\delta)}{\partial\sigma^2} + \frac{\partial L_n(\theta_{10},\delta)}{\partial\delta} = O(\delta^2)$, where $\theta_{10} = (\beta_0', \sigma_0^2)'$ and $\psi_0 = \phi(0)/[1 - \Phi(0)]$. Thus, when δ_0 is nonzero but small, the MLE and thus the PMLEs can perform poorly, which can be seen from Table 7. When the sample size increases from 200 to 600, the probabilities for $\delta_0 = 2$ and $\delta_0 = 0$ increase, but others decrease except that of the PMLE-o with $\delta_0 = 1$.

Table 6. Probabilities that the PMLEs of the stochastic frontier function model select the right model.

	$n = 200$		$n = 600$	
	PMLE-o	**PMLE-t**	**PMLE-o**	**PMLE-t**
$\delta_0 = 2$	0.822	0.838	0.991	0.991
$\delta_0 = 1$	0.170	0.289	0.196	0.271
$\delta_0 = 0.5$	0.071	0.184	0.025	0.082
$\delta_0 = 0.25$	0.054	0.132	0.012	0.065
$\delta_0 = 0.1$	0.050	0.159	0.015	0.059
$\delta_0 = 0$	0.961	0.856	0.990	0.925

The PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. When $\delta_0 \neq 0$, the numbers in the table are the probabilities that the PMLEs of δ are non-zero; when $\delta_0 = 0$, the numbers are the probabilities that the PMLEs of δ are zero.

Table 7 presents biases, SEs and RMSEs of the MLE, PMLE-o, PMLE-t and MLE-r with the restriction $\delta = 0$ imposed, even though $\delta_0 \neq 0$. Since the MLE-r imposes the wrong restriction, it has

very large biases for β_1, σ^2 and δ but it generally has the smallest SEs. The MLE, PMLE-o and PMLE-t of β_2 and β_3 have similar features. For $\delta_0 = 2$, 1 and 0.5, the biases of the PMLEs of β_1, σ^2 and δ are generally larger than those of the MLE, but are smaller than those of the MLE-r. The SEs of the PMLEs are larger than those of the MLE for $\delta_0 = 2$ and 1 but are smaller for smaller values of δ_0. For $\delta_0 = 0.25$ and 0.1, even though the PMLEs estimate δ as zero with high probabilities, they have smaller biases, SEs and RMSEs than those of the MLE in almost all cases. As the sample size n increases, all estimates have smaller SEs, the MLEs have smaller biases, but the MLE-r and PMLEs may have smaller or larger biases.

Table 7. The biases, SEs and RMSEs of the estimators when $\delta_0 \neq 0$ in the stochastic frontier function model.

n, δ_0		β_1	β_2	β_3	σ^2	δ
200, 2	MLE-r	−1.595[0.112]1.599	0.002[0.114]0.114	−0.001[0.057]0.057	−2.574[0.264]2.588	−2.000[0.000]2.000
	MLE	−0.034[0.301]0.303	0.002[0.110]0.110	−0.002[0.055]0.055	−0.050[0.996]0.998	0.115[0.724]0.733
	PMLE-o	−0.235[0.662]0.703	0.002[0.111]0.111	−0.002[0.056]0.056	−0.291[1.348]1.379	−0.093[1.047]1.051
	PMLE-t	−0.215[0.640]0.675	0.002[0.111]0.111	−0.002[0.055]0.055	−0.266[1.319]1.345	−0.072[1.021]1.024
200, 1	MLE-r	−0.795[0.082]0.799	0.002[0.082]0.082	0.001[0.041]0.041	−0.657[0.134]0.671	−1.000[0.000]1.000
	MLE	−0.136[0.426]0.447	0.002[0.082]0.082	0.001[0.042]0.042	−0.050[0.522]0.524	−0.077[0.657]0.661
	PMLE-o	−0.602[0.438]0.744	0.002[0.082]0.082	0.001[0.042]0.042	−0.434[0.536]0.690	−0.684[0.713]0.988
	PMLE-t	−0.499[0.484]0.695	0.002[0.082]0.082	0.001[0.042]0.042	−0.343[0.561]0.657	−0.546[0.756]0.932
200, 0.5	MLE-r	−0.395[0.073]0.401	0.002[0.070]0.070	0.000[0.039]0.039	−0.178[0.106]0.207	−0.500[0.000]0.500
	MLE	−0.014[0.380]0.380	0.002[0.071]0.071	0.000[0.039]0.039	0.106[0.363]0.378	0.068[0.600]0.604
	PMLE-o	−0.324[0.267]0.420	0.002[0.071]0.071	0.000[0.039]0.039	−0.107[0.284]0.304	−0.373[0.470]0.600
	PMLE-t	−0.242[0.341]0.418	0.002[0.071]0.071	0.000[0.039]0.039	−0.045[0.326]0.330	−0.251[0.559]0.613
200, 0.25	MLE-r	−0.199[0.071]0.211	−0.003[0.071]0.071	−0.001[0.034]0.034	−0.052[0.102]0.115	−0.250[0.000]0.250
	MLE	0.120[0.362]0.382	−0.003[0.071]0.071	−0.002[0.034]0.034	0.177[0.329]0.373	0.235[0.572]0.618
	PMLE-o	−0.147[0.232]0.275	−0.003[0.071]0.071	−0.002[0.034]0.034	−0.002[0.244]0.244	−0.158[0.389]0.420
	PMLE-t	−0.093[0.288]0.302	−0.003[0.071]0.071	−0.002[0.034]0.034	0.037[0.271]0.273	−0.075[0.472]0.478
200, 0.1	MLE-r	−0.079[0.073]0.108	−0.002[0.071]0.071	0.002[0.037]0.037	−0.018[0.105]0.107	−0.100[0.000]0.100
	MLE	0.240[0.355]0.429	−0.002[0.071]0.071	0.002[0.037]0.037	0.208[0.314]0.377	0.391[0.573]0.694
	PMLE-o	−0.032[0.214]0.216	−0.002[0.071]0.071	0.002[0.037]0.037	0.027[0.229]0.231	−0.013[0.384]0.384
	PMLE-t	0.046[0.296]0.299	−0.002[0.071]0.071	0.002[0.037]0.037	0.085[0.278]0.291	0.108[0.503]0.514
600, 2	MLE-r	−1.595[0.066]1.596	−0.004[0.065]0.066	0.001[0.033]0.033	−2.558[0.151]2.563	−2.000[0.000]2.000
	MLE	−0.007[0.142]0.142	−0.003[0.061]0.061	0.000[0.031]0.031	−0.016[0.540]0.541	0.038[0.349]0.351
	PMLE-o	−0.017[0.204]0.204	−0.004[0.061]0.061	0.000[0.031]0.031	−0.028[0.582]0.583	0.028[0.390]0.391
	PMLE-t	−0.017[0.204]0.204	−0.004[0.061]0.061	0.000[0.031]0.031	−0.028[0.582]0.583	0.028[0.390]0.391
600, 1	MLE-r	−0.796[0.047]0.797	0.004[0.048]0.049	0.001[0.025]0.025	−0.640[0.079]0.645	−1.000[0.000]1.000
	MLE	−0.073[0.288]0.297	0.004[0.048]0.048	0.000[0.025]0.025	−0.036[0.350]0.352	−0.062[0.417]0.422
	PMLE-o	−0.597[0.406]0.722	0.004[0.048]0.048	0.000[0.025]0.025	−0.438[0.431]0.614	−0.717[0.577]0.921
	PMLE-t	−0.536[0.433]0.689	0.004[0.048]0.048	0.000[0.025]0.025	−0.387[0.445]0.590	−0.639[0.605]0.880
600, 0.5	MLE-r	−0.397[0.042]0.399	−0.002[0.043]0.043	−0.000[0.022]0.022	−0.165[0.063]0.176	−0.500[0.000]0.500
	MLE	−0.062[0.316]0.322	−0.002[0.043]0.043	−0.000[0.022]0.022	0.047[0.248]0.252	−0.040[0.449]0.451
	PMLE-o	−0.375[0.142]0.401	−0.002[0.043]0.043	−0.000[0.022]0.022	−0.145[0.141]0.202	−0.466[0.215]0.513
	PMLE-t	−0.336[0.210]0.396	−0.002[0.043]0.043	−0.000[0.022]0.022	−0.118[0.177]0.212	−0.410[0.309]0.513
600, 0.25	MLE-r	−0.200[0.041]0.204	0.001[0.042]0.042	0.001[0.021]0.021	−0.046[0.059]0.075	−0.250[0.000]0.250
	MLE	0.065[0.289]0.296	0.001[0.042]0.042	0.001[0.021]0.021	0.107[0.202]0.229	0.121[0.414]0.432
	PMLE-o	−0.190[0.101]0.215	0.001[0.042]0.042	0.001[0.021]0.021	−0.037[0.100]0.107	−0.234[0.149]0.277
	PMLE-t	−0.158[0.170]0.232	0.001[0.042]0.042	0.001[0.021]0.021	−0.017[0.131]0.132	−0.187[0.247]0.310
600, 0.1	MLE-r	−0.080[0.040]0.089	−0.003[0.041]0.041	−0.001[0.020]0.020	−0.011[0.058]0.059	−0.100[0.000]0.100
	MLE	0.187[0.295]0.350	−0.003[0.041]0.041	−0.001[0.020]0.020	0.145[0.205]0.251	0.279[0.427]0.510
	PMLE-o	−0.067[0.110]0.129	−0.003[0.041]0.041	−0.001[0.020]0.020	−0.000[0.100]0.100	−0.079[0.169]0.187
	PMLE-t	−0.039[0.172]0.176	−0.003[0.041]0.041	−0.001[0.020]0.020	0.018[0.132]0.133	−0.037[0.258]0.261

The MLE-r denotes the restricted MLE with the restriction $\delta = 0$ imposed, and PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias[SE]RMSE. $\beta_0 = (1, 1, 1)'$. Corresponding to $\delta_0 = 2$, 1, 0.5, 0.25 and 0.1, the true value of σ^2 is $\sigma_0^2 = 5$, 2, 1.25, 1.0625 and 1.01.

The biases, SEs and RMSEs of the estimators when $\delta_0 = 0$ are presented in Table 8. All estimators of various estimation methods have similar summary statistics for β_2 and β_3. For other parameters, the MLE-r has the smallest biases, SEs and RMSEs, since it imposes the correct restriction $\delta = 0$. The PMLEs have much smaller biases, SEs and RMSEs than those of the MLE. The biases, SEs and RMSEs of the PMLE-o are smaller than those of the PMLE-t. As the sample size increases to 600, the summary statistics of the PMLE-o become very close to those of the MLE-r. For all estimates, we observe smaller biases, SEs and RMSEs for a larger sample size.

Table 8. The biases, SEs and RMSEs of the estimators when $\delta_0 = 0$ in the stochastic frontier function model.

n, δ_0		β_1	β_2	β_3	σ^2	δ
200, 0	MLE-r	−0.000[0.074]0.074	−0.001[0.073]0.073	−0.001[0.037]0.037	−0.016[0.100]0.101	0.000[0.000]0.000
	MLE	0.302[0.347]0.460	−0.001[0.073]0.073	−0.002[0.037]0.037	0.191[0.295]0.351	0.462[0.549]0.718
	PMLE-o	0.037[0.198]0.202	−0.001[0.073]0.073	−0.002[0.037]0.037	0.018[0.202]0.203	0.067[0.337]0.344
	PMLE-t	0.109[0.278]0.298	−0.001[0.073]0.073	−0.002[0.037]0.037	0.069[0.248]0.257	0.178[0.459]0.492
600, 0	MLE-r	0.001[0.040]0.040	−0.001[0.041]0.042	−0.001[0.022]0.022	−0.002[0.057]0.057	0.000[0.000]0.000
	MLE	0.268[0.292]0.396	−0.001[0.042]0.042	−0.001[0.022]0.022	0.153[0.206]0.257	0.377[0.419]0.564
	PMLE-o	0.009[0.093]0.093	−0.001[0.042]0.042	−0.001[0.022]0.022	0.005[0.089]0.089	0.014[0.138]0.139
	PMLE-t	0.049[0.178]0.185	−0.001[0.042]0.042	−0.001[0.022]0.022	0.031[0.132]0.135	0.072[0.262]0.272

The MLE-r denotes the restricted MLE with the restriction $\delta = 0$ imposed, and PMLE-o and PMLE-t denote the PMLEs obtained from the criterion functions formulated using, respectively, the original and transformed likelihood functions. The three numbers in each cell are bias[SE]RMSE. $\beta_0 = (1, 1, 1)'$ and $\sigma_0^2 = 1$.

5. Conclusions

In this paper, we investigate the estimation of parametric models with singular information matrices using the PML based on the adaptive lasso (group lasso). An irregular model has a singular information matrix occurring at a subvector θ_{20} of the true parameter vector θ_0 being zero, but its information matrices at other parameter values are nonsingular. In addition, if we knew that θ_{20} is zero, the restricted model always has a nonsingular information matrix. We show that the PMLEs have oracle properties. Consequently, the PMLEs always have the $\sqrt{n}$-rate of convergence, no matter whether $\theta_{20} = 0$ or not, while the MLEs usually have slower than the $\sqrt{n}$-rate of convergence and their asymptotic distributions might not be normal when $\theta_{20} = 0$. The PML can conduct model selection and estimation simultaneously. As examples, we consider the PMLEs for the sample selection model and the stochastic frontier function model, which can be formulated with both original structural parameters of interest and transformed parameters. Our Monte Carlo results show that the PMLE formulated with the original parameters generally performs well and outperforms the reparameterized one in terms of smaller RMSEs.

Acknowledgments: Fei Jin gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71501119).

Author Contributions: The authors have contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. MLE of the Sample Selection Model

In this section, we derive the asymptotic distribution of the MLE of the sample selection model (3). The irregularity of the information matrix occurs at $\rho_0 = 0$, which is in the interior of the range for the correlation coefficient. So for this model, the true parameter vector θ_0 of interest is in the interior of the compact parameter space Θ. In addition, we assume that the exogenous variables x_i and z_i are uniformly bounded, the empirical distribution of (x_i, z_i) converges in distribution to a limiting distribution and the matrices $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i x_i'$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n z_i z_i'$ exist and are positive definite. These assumptions are strong enough to establish the asymptotic properties in this section.

The log likelihood function of model (3) divided by n is

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^n \left\{ (1 - I_i)\ln(1 - \Phi(z_i'\gamma)) - \frac{1}{2} I_i \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} I_i (y_i - x_i'\beta)^2 + I_i \ln \Phi\left[\frac{1}{\sqrt{1-\rho^2}} \left(z_i'\gamma - \frac{\rho(y_i - x_i'\beta)}{\sigma} \right) \right] \right\}, \quad \text{(A1)}$$

where $\theta = (\beta', \sigma^2, \gamma', \rho)'$ and $\Phi(\cdot)$ is the standard normal distribution. The first order derivatives of $L_n(\theta)$ are

$$\frac{\partial L_n(\theta)}{\partial \beta} = \frac{1}{n\sigma^2}\sum_{i=1}^{n} I_i x_i[\epsilon_i(\beta) + \sigma\rho(1-\rho^2)^{-1/2}\psi_i(\theta)], \tag{A2}$$

$$\frac{\partial L_n(\theta)}{\partial \sigma^2} = \frac{1}{2n\sigma^2}\sum_{i=1}^{n} I_i\big[\frac{\epsilon_i^2(\beta)}{\sigma^2} - 1 + \frac{1}{\sigma}\rho(1-\rho^2)^{-1/2}\psi_i(\theta)\epsilon_i(\beta)\big], \tag{A3}$$

$$\frac{\partial L_n(\theta)}{\partial \gamma} = \frac{1}{n}\sum_{i=1}^{n} z_i\big[(1-\rho^2)^{-1/2} I_i\psi_i(\theta) - (1-I_i)\frac{\phi_i}{1-\Phi_i}\big], \tag{A4}$$

$$\frac{\partial L_n(\theta)}{\partial \rho} = \frac{1}{n}(1-\rho^2)^{-3/2}\sum_{i=1}^{n} I_i\psi_i(\theta)\big(\rho z_i'\gamma - \frac{\epsilon_i(\beta)}{\sigma}\big), \tag{A5}$$

where $\epsilon_i(\beta) = y_i - x_i'\beta$, $\phi_i = \phi(z_i'\gamma)$, $\Phi_i = \Phi(z_i'\gamma)$ and $\psi_i(\theta) = \phi\big((1-\rho^2)^{-1/2}(z_i'\gamma - \frac{\rho}{\sigma}\epsilon_i(\beta))\big)/\Phi\big((1-\rho^2)^{-1/2}(z_i'\gamma - \frac{\rho}{\sigma}\epsilon_i(\beta))\big)$ with $\phi(\cdot)$ being the standard normal pdf. It is known that the variance-covariance matrix of a vector of random variables is positive definite if and only if there is no linear relation among the components of the random vector (Rao 1973, p. 107). Under the assumed regularity conditions, one can easily show that when $\rho_0 \neq 0$, the gradients (A2)–(A5) at θ_0 are linearly independent w.p.a.1., and hence the limiting matrix of $\frac{1}{n}I_n(\theta_0)$, where $I_n(\theta_0)$ is the information matrix with the sample size n, is positive definite. Thus, there are no irregularities in the model when $\rho_0 \neq 0$, and the MLE is $\sqrt{n}$-consistent and asymptotically normal.

However, when $\rho_0 = 0$ and together with γ_0, there are some irregularities in the model. With $\rho_0 = 0$, the first order derivatives are

$$\frac{\partial L_n(\theta_0)}{\partial \beta} = \frac{1}{n\sigma_0^2}\sum_{i=1}^{n} I_i x_i\epsilon_i, \tag{A6}$$

$$\frac{\partial L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2n\sigma_0^4}\sum_{i=1}^{n} I_i(\epsilon_i^2 - \sigma_0^2), \tag{A7}$$

$$\frac{\partial L_n(\theta_0)}{\partial \gamma} = \frac{1}{n}\sum_{i=1}^{n} \frac{[I_i - \Phi(z_i'\gamma_0)]\phi(z_i'\gamma_0)}{\Phi(z_i'\gamma_0)[1-\Phi(z_i'\gamma_0)]} z_i, \tag{A8}$$

$$\frac{\partial L_n(\theta_0)}{\partial \rho} = -\frac{1}{n\sigma_0}\sum_{i=1}^{n} \frac{\phi(z_i'\gamma_0)}{\Phi(z_i'\gamma_0)} I_i\epsilon_i. \tag{A9}$$

These derivatives are linearly independent as long as x and $\phi(z'\gamma_0)/\Phi(z'\gamma_0)$ are linearly independent, which will usually be the case if z contains some relevant continuous exogenous variables with nonzero coefficients. However, when the non-intercept variables in z have coefficients equal to zero, $\phi(z'\gamma_0)/\Phi(z'\gamma_0)$ is a constant for all i, and the first component of $\frac{\partial L_n(\alpha_0,0)}{\partial \beta}$ and $\frac{\partial L_n(\alpha_0,0)}{\partial \rho}$ are linearly dependent as x contains an intercept term. It follows that the information matrix must be singular. We consider this irregularity below. Let $x_i = (1, x_{2i}')'$, $\beta = (\beta_1, \beta_2')'$ with β_1 being a scalar, $\gamma = (\gamma_1, \gamma_2')'$ with γ_1 being the coefficient for the intercept term of the selection equation, $\alpha = (\beta', \sigma^2, \gamma_1)'$, $\theta_2 = (\gamma_2', \rho)'$, $\theta = (\alpha', \theta_2')'$, and $\theta_{20} = 0$. Then,

$$\frac{\partial L_n(\theta_0)}{\partial \rho} + \sigma_0\psi_0\frac{\partial L_n(\theta_0)}{\partial \beta_1} = 0. \tag{A10}$$

where $\psi_0 = \phi(\gamma_{10})/\Phi(\gamma_{10})$. Furthermore, the submatrix of the information matrix corresponding to α with the sample size n is

$$\Xi_n = \mathrm{E}\big(n^2\frac{\partial L_n(\theta_0)}{\partial \alpha}\frac{\partial L_n(\theta_0)}{\partial \alpha'}\big) = \begin{pmatrix} \frac{\Phi(\gamma_0)}{\sigma_0^2}\sum_{i=1}^{n} x_i x_i' & 0 & 0 \\ 0 & \frac{n\Phi(\gamma_0)}{2\sigma_0^4} & 0 \\ 0 & 0 & \frac{n\phi(\gamma_{10})^2}{\Phi(\gamma_{10})[1-\Phi(\gamma_{10})]} \end{pmatrix}.$$

The limit of Ξ_n/n has full rank under the assumed regularity conditions. Thus, the rank of the information matrix is one less than the total number of parameters. This sample selection model (3) has irregularities similar to the stochastic frontier function model in Lee (1993), with the exception that the true parameter vector is not on the boundary of a parameter space. The asymptotic distribution of its MLE can be similarly derived. The method in Rotnitzky et al. (2000) can also be used, but the method in Lee (1993) is simpler for this particular model.

Consider the transformation of $(\alpha', \theta_2)'$ to $(\xi', \theta_2)'$ defined by $\xi = \alpha - \rho K_1$, where $K_1 = (\sigma_0\psi_0, 0_{1\times(k_x+1)})'$ with k_x being the number of variables in x.[9] At $\rho_0 = 0$, $\xi_0 = \alpha_0$. Define $L_{n1}(\xi, \rho)$ by

$$L_{n1}(\xi, \theta_2) = L_n(\xi + \rho K_1, \theta_2), \tag{A11}$$

which is the log likelihood divided by n in terms of ξ and θ_2. Then

$$\frac{\partial L_{n1}(\xi_0, 0)}{\partial \xi} = \frac{\partial L_n(\alpha_0, 0)}{\partial \alpha}, \tag{A12}$$

and by (A10),

$$\frac{\partial L_{n1}(\xi_0, 0)}{\partial \rho} = \frac{\partial L_n(\alpha_0, 0)}{\partial \rho} + \sigma_0\psi_0 \frac{\partial L_n(\alpha_0, 0)}{\partial \beta_1} = 0. \tag{A13}$$

Thus, the derivative of $L_{n1}(\xi, \theta_2)$ with respect to ρ at $(\xi_0', 0)'$ is zero. The derivative can be interpreted as the residual vector $\frac{\partial L_n(\alpha_0,0)}{\partial \rho} - \left[\mathrm{E}\left(\frac{\partial L_n(\alpha_0,0)}{\partial \rho}\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}\right)\right]\left[\mathrm{E}\left(\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}\right)^2\right]^{-1}\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}$ of the minimum mean square regression of $\frac{\partial L_n(\alpha_0,0)}{\partial \rho}$ on $\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}$. The linear dependence relation (A10) implies that the residual vector must be zero and $\left[\mathrm{E}\left(\frac{\partial L_n(\alpha_0,0)}{\partial \rho}\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}\right)\right]\left[\mathrm{E}\left(\frac{\partial L_n(\alpha_0,0)}{\partial \beta_1}\right)^2\right]^{-1} = -\sigma_0\psi_0$. Furthermore, we see that

$$\begin{aligned}\frac{\partial^2 L_{n1}(\xi_0, 0)}{\partial \rho^2} &= \frac{\partial^2 L_n(\alpha_0, 0)}{\partial \rho^2} + 2\sigma_0\psi_0 \frac{\partial^2 L_n(\alpha_0, 0)}{\partial \rho \partial \beta_1} + \sigma_0^2\psi_0^2 \frac{\partial^2 L_n(\alpha_0, 0)}{\partial \beta_1^2} \\ &= \frac{\psi_0(\psi_0 + \gamma_{10})}{n\sigma_0^2} \sum_{i=1}^{n} I_i(\sigma_0^2 - \epsilon_i^2).\end{aligned}$$

Then by (A12) and (A7),

$$\frac{\partial^2 L_{n1}(\xi_0, 0)}{\partial \rho^2} + 2\sigma_0^2\psi_0(\psi_0 + \gamma_{10}) \frac{\partial L_{n1}(\xi_0, 0)}{\partial \xi_{k_x+1}} = 0, \tag{A14}$$

where ξ_{k_x+1} denotes the $(k_x + 1)$th component of ξ. This is a second irregularity of the model. Following Lee (1993) and Rotnitzky et al. (2000), consider the transformation of $(\kappa', \rho)'$ to $(\eta', \rho)'$ defined by $\eta = \kappa - \frac{1}{2}\rho^2 K_2$, where $\kappa = (\xi', \gamma_2')'$ and $K_2 = [0_{1\times k_x}, 2\sigma_0^2\psi_0(\psi_0 + \gamma_{10}), 0_{1\times(k_z+1)}]'$ with k_z being the number of parameters in z, and the function $L_{n2}(\eta, \rho)$ defined by

$$L_{n2}(\eta, \rho) = L_{n1}(\eta + \frac{1}{2}\rho^2 K_2, \rho). \tag{A15}$$

---

[9] For the reparameterization in Lee (1993), the parameters σ and ψ in K_1 are not taken to be the true values. Both methods work. The method here might be simpler in computation.

Then

$$\frac{\partial L_{n2}(\eta,\rho)}{\partial \eta} = \frac{\partial L_{n1}(\kappa,\rho)}{\partial \kappa}, \tag{A16}$$

$$\frac{\partial L_{n2}(\eta,\rho)}{\partial \rho} = \rho \frac{\partial L_{n1}(\kappa,\rho)}{\partial \kappa'} K_2 + \frac{\partial L_{n1}(\kappa,\rho)}{\partial \rho}, \tag{A17}$$

$$\frac{\partial^2 L_{n2}(\eta,\rho)}{\partial \rho^2} = \rho^2 K_2' \frac{\partial^2 L_{n1}(\kappa,\rho)}{\partial \kappa \partial \kappa'} K_2 + 2\rho \frac{\partial^2 L_{n1}(\kappa,\rho)}{\partial \rho \partial \kappa'} K_2 + \frac{\partial L_{n1}(\kappa,\rho)}{\partial \kappa'} K_2 + \frac{\partial^2 L_{n1}(\kappa,\rho)}{\partial \rho^2}. \tag{A18}$$

At $\rho_0 = 0$, $\eta_0 = \kappa_0$. By (A13) and the linear dependence relation in (A14),

$$\frac{\partial L_{n2}(\eta_0,0)}{\partial \eta} = \frac{\partial L_{n1}(\kappa_0,0)}{\partial \kappa}, \tag{A19}$$

$$\frac{\partial L_{n2}(\eta_0,0)}{\partial \rho} = 0, \tag{A20}$$

and

$$\frac{\partial^2 L_{n2}(\eta_0,0)}{\partial \rho^2} = 0. \tag{A21}$$

Since the first and second order derivatives of $L_{n2}(\eta,\rho)$ with respect to ρ at $(\eta_0,0)$ are zero, it is necessary to investigate the third order derivative of $L_{n2}(\eta,\rho)$ with respect to ρ at $(\eta_0,0)$. By (A18) and (A10),

$$\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3} = 3\frac{\partial^2 L_{n1}(\kappa_0,0)}{\partial \rho \partial \kappa'} K_2 + \frac{\partial^3 L_{n1}(\kappa_0,0)}{\partial \rho^3}. \tag{A22}$$

Note that $3\frac{\partial^2 L_{n1}(\kappa_0,0)}{\partial \rho \partial \kappa'} K_2 = 6\sigma_0^2\psi_0(\psi_0+\gamma_{10})\frac{\partial^2 L_{n1}(\kappa_0,0)}{\partial \rho \partial \kappa_{k_x+1}}$. Since $\frac{\partial L_{n1}(\kappa,\rho)}{\partial \rho} = \sigma_0\psi_0\frac{\partial L_n(\alpha,\rho)}{\partial \beta_1} + \frac{\partial L_n(\alpha,\rho)}{\partial \rho}$, $\frac{\partial^2 L_{n1}(\kappa,\rho)}{\partial \rho \partial \kappa_{k_x+1}} = \sigma_0\psi_0\frac{\partial^2 L_n(\alpha,\rho)}{\partial \beta_1 \partial \sigma^2} + \frac{\partial^2 L_n(\alpha,\rho)}{\partial \rho \partial \sigma^2}$, $\frac{\partial^2 L_{n1}(\kappa,\rho)}{\partial \rho^2} = \sigma_0^2\psi_0^2\frac{\partial^2 L_n(\alpha,\rho)}{\partial \beta_1^2} + 2\sigma_0\psi_0\frac{\partial^2 L_n(\alpha,\rho)}{\partial \beta_1 \partial \rho} + \frac{\partial^2 L_n(\alpha,\rho)}{\partial \rho^2}$, and $\frac{\partial^3 L_{n1}(\kappa,\rho)}{\partial \rho^3} = \sigma_0^3\psi_0^3\frac{\partial^3 L_n(\alpha,\rho)}{\partial \beta_1^3} + 3\sigma_0^2\psi_0^2\frac{\partial^3 L_n(\alpha,\rho)}{\partial \beta_1^2 \partial \rho} + 3\sigma_0\psi_0\frac{\partial^3 L_n(\alpha,\rho)}{\partial \beta_1 \partial \rho^2} + \frac{\partial^3 L_n(\alpha,\rho)}{\partial \rho^3}$, it is straightforward to show that

$$3\frac{\partial^2 L_{n1}(\kappa_0,0)}{\partial \rho \partial \kappa'} K_2 = -\frac{3}{n}\psi_0^2(\psi_0+\gamma_{10})\sum_{i=1}^{n} I_i\left(\frac{\epsilon_i}{\sigma_0}\right),$$

and

$$\frac{\partial^3 L_{n1}(\kappa_0,0)}{\partial \rho^3} = \frac{1}{n}\psi_0(1-2\psi_0^2-3\psi_0\gamma_{10}-\gamma_{10}^2)\sum_{i=1}^{n} I_i\left[\left(\frac{\epsilon_i}{\sigma_0}\right)^3 - 3\left(\frac{\epsilon_i}{\sigma_0}\right)\right].$$

Then

$$\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3} = -\frac{3}{n}\psi_0^2(\psi_0+\gamma_{10})\sum_{i=1}^{n} I_i\left(\frac{\epsilon_i}{\sigma_0}\right) + \frac{1}{n}\psi_0(1-2\psi_0^2-3\psi_0\gamma_{10}-\gamma_{10}^2)\sum_{i=1}^{n} I_i\left[\left(\frac{\epsilon_i}{\sigma_0}\right)^3 - 3\left(\frac{\epsilon_i}{\sigma_0}\right)\right]. \tag{A23}$$

Thus, $\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3}$ is not linearly dependent on $\frac{\partial L_{n2}(\eta_0,0)}{\partial \eta}$. Under this circumstance, as in Rotnitzky et al. (2000), the asymptotic distribution of the MLE can be derived by investigating high order Taylor expansions of the first order condition of $L_{n2}(\eta,\rho)$. For the stochastic frontier function model, Lee (1993) shows that the asymptotic distribution of the MLE can be derived by considering

one more reparameterization. We employ the approach in Lee (1993).[10] Note that a Taylor expansion of $\frac{\partial L_{n2}(\eta_0,\rho)}{\partial \rho}$ around $\rho = 0$ up to the second order yields

$$\frac{\partial L_{n2}(\eta_0,\rho)}{\partial \rho} = \frac{\partial L_{n2}(\eta_0,0)}{\partial \rho} + \frac{\partial^2 L_{n2}(\eta_0,0)}{\partial \rho^2}\rho + \frac{1}{2}\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3}\rho^2 + o(\rho^2) = \frac{1}{2}\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3}\rho^2 + o(\rho^2),$$

where the second equality follows by (A20) and (A21). Consider the transformation of (η, δ) to (η, r) defined by

$$r = \rho^3, \tag{A24}$$

and the function $L_{n3}(\eta, r)$ defined by

$$L_{n3}(\eta, r) = L_{n2}(\eta, r^{1/3}). \tag{A25}$$

It follows that

$$\frac{\partial L_{n3}(\eta,r)}{\partial \eta} = \frac{\partial L_{n2}(\eta,\delta)}{\partial \eta}, \text{ and } \frac{\partial L_{n3}(\eta,r)}{\partial r} = \frac{1}{3\rho^2}\frac{\partial L_{n2}(\eta,r)}{\partial \rho}. \tag{A26}$$

Hence,

$$\frac{\partial L_{n3}(\eta_0,0)}{\partial \eta} = \frac{\partial L_{n2}(\eta_0,0)}{\partial \eta}, \text{ and } \frac{\partial L_{n3}(\eta_0,0)}{\partial r} = \frac{1}{6}\frac{\partial^3 L_{n2}(\eta_0,0)}{\partial \rho^3}. \tag{A27}$$

From (A27) and (A23), $\frac{\partial L_{n3}(\eta_0,0)}{\partial \eta}$ and $\frac{\partial L_{n3}(\eta_0,0)}{\partial r}$ are linearly independent. Then the information matrix for $L_{n3}(\eta, r)$ is nonsingular and the MLE $(\tilde{\eta}_n', \tilde{r}_n)'$ has the asymptotic distribution

$$\sqrt{n}(\tilde{\eta}_n' - \eta_0', \tilde{r}_n)' \xrightarrow{d} N(0, \lim_{n\to\infty} \Omega_n), \tag{A28}$$

where

$$\Omega_n = \begin{pmatrix} \frac{\Phi(\gamma_{10})}{n\sigma_0^2}\sum_{i=1}^n x_i x_i' & 0 & 0 & -\frac{\psi_0^2(\psi_0+\gamma_{10})\Phi(\gamma_0)}{2n\sigma_0}\sum_{i=1}^n x_i \\ 0 & \frac{\Phi(\gamma_{10})}{2\sigma_0^4} & 0 & 0 \\ 0 & 0 & \frac{\phi^2(\gamma_{10})}{n\Phi(\gamma_{10})[1-\Phi(\gamma_{10})]}\sum_{i=1}^n z_i z_i' & 0 \\ -\frac{\psi_0^2(\psi_0+\gamma_{10})\Phi(\gamma_{10})}{2n\sigma_0}\sum_{i=1}^n x_i' & 0 & 0 & \frac{1}{12}\Phi(\gamma_{10})[3\psi_0^4(\psi_0+\gamma_{10})^2 + 2\psi_0^2(1-2\psi_0^2-3\psi_0\gamma_{10}-\gamma_{10}^2)^2] \end{pmatrix}. \tag{A29}$$

The complete transformation for the model is

$$\eta_1 = \beta_1 - \sigma_0\psi_0\rho, \quad \eta_2 = \beta_2, \quad \eta_3 = \sigma^2 - \rho^2\sigma_0^2\psi_0(\psi_0+\gamma_{10}), \quad \eta_4 = \gamma, \quad r = \rho^3.$$

The inverse transformation is

$$\beta_1 = \eta_1 + \sigma_0\psi_0 r^{1/3}, \tag{A30}$$

$$\beta_2 = \eta_2, \tag{A31}$$

$$\sigma^2 = \eta_3 + r^{2/3}\sigma_0^2\psi_0(\psi_0+\gamma_{10}), \tag{A32}$$

$$\gamma = \eta_4, \tag{A33}$$

$$\rho = r^{1/3}. \tag{A34}$$

[10] In Rotnitzky et al. (2000), for a general model, it is possible that the order of the first non-zero derivative with respect to the first component (last component in this paper) is either odd or even after proper reparameterizations. If the order is even, there is a need to analyze the sign of the MLE. In our case, the order is odd and the asymptotic distribution of the MLE can be derived by considering one more reparameterization.

With the asymptotic distribution of $(\tilde{\eta}_n', \tilde{\rho}_n)'$ in (A28), the asymptotic distribution of the MLE $(\tilde{\beta}_n', \tilde{\sigma}_n^2, \tilde{\gamma}_n, \tilde{\rho}_n)'$ for the original parameters can then be derived from the inverse transformations (A30)–(A34) by Slutsky's theorem and the continuous mapping theorem. From (A34), $\tilde{\rho}_n = \tilde{r}_n^{1/3}$. By the matrix inverse formula in a block form, $\sqrt{n}\tilde{r}_n$ is asymptotically normal $N\big(0, \frac{1}{6}\Phi(\gamma_0)\psi_0^2(1-2\psi_0^2-3\psi_0\gamma_{10}-\gamma_{10}^2)^2\big)$. Then it follows that $n^{1/6}\tilde{\rho}_n = (n^{1/2}\tilde{r}_n)^{1/3}$ is asymptotically distributed as a cubic root of a normal variable, and $\tilde{\rho}_n$ converges in distribution at a much lower rate of convergence.[11] Since

$$n^{1/6}(\tilde{\beta}_{1n} - \beta_{10}) = n^{1/6}(\tilde{\eta}_{1n} - \eta_{10}) + \sigma_0\psi_0(n^{1/2}\tilde{r}_n)^{1/3} = \sigma_0\psi_0(n^{1/2}\tilde{r}_n)^{1/3} + o_p(1),$$

the MLE $\tilde{\beta}_{1n}$ has the same rate of convergence as $\tilde{\rho}_n$, and the asymptotic distribution of $n^{1/6}(\tilde{\beta}_{1n} - \beta_{10})$ is the same as that of $\sigma_0\psi_0(n^{1/2}\tilde{r}_n)^{1/3}$. Similarly, as

$$n^{1/3}(\tilde{\sigma}_n^2 - \sigma_0^2) = n^{1/3}(\tilde{\eta}_{3n} - \eta_{30}) + \sigma_0^2\psi_0(\psi_0 + \gamma_{10})(n^{1/2}\tilde{r}_n)^{2/3} = \sigma_0^2\psi_0(\psi_0 + \gamma_{10})(n^{1/2}\tilde{r}_n)^{2/3} + o_p(1),$$

$n^{1/3}(\tilde{\sigma}_n^2 - \sigma_0^2)$ has the same asymptotic distribution as $\sigma_0^2\psi_0(\psi_0 + \gamma_{10})(n^{1/2}\tilde{r}_n)^{2/3}$. Both $\tilde{\beta}_{1n}$ and $\tilde{\sigma}_n^2$ converge in distribution at some lower rates of convergence and are not asymptotically normally distributed. $n^{1/6}(\tilde{\beta}_{1n} - \beta_{10})$ is asymptotically distributed as a cubic root of a normal variable and is asymptotically proportional to $n^{1/6}\tilde{\rho}_n$. $n^{1/3}(\tilde{\sigma}_n^2 - \sigma_0^2)$ is asymptotically distributed as a 2/3 power of a normal variable. The remaining estimates $\tilde{\beta}_{2n}$ and $\tilde{\gamma}_n$, however, have the usual order $O_p(n^{-1/2})$ and $\sqrt{n}\left(\begin{smallmatrix}\tilde{\beta}_{2n}-\beta_{20}\\ \tilde{\gamma}_n-\gamma_0\end{smallmatrix}\right) = \sqrt{n}\left(\begin{smallmatrix}\tilde{\eta}_{2n}-\eta_{20}\\ \tilde{\eta}_{4n}-\eta_{40}\end{smallmatrix}\right)$ is asymptotically normally distributed. From the information matrix in (A29), the joint asymptotic distribution of $\tilde{\beta}_n$, $\tilde{\sigma}_n^2$, $\tilde{\gamma}_n$, and $\tilde{\rho}_n$ can also be derived.

Appendix B. Proofs

Proof of Proposition 1. When $\theta_{20} \neq 0$, by Assumption 2, $\tilde{\theta}_{2n} = \theta_{20} + o_p(1)$ and $\|\tilde{\theta}_{2n}\|^{-\mu} = O_p(1)$. Then, w.p.a.1.,

$$Q_n(\hat{\theta}_n) = L_n(\hat{\theta}_n) - \lambda_n\|\tilde{\theta}_{2n}\|^{-\mu}\|\hat{\theta}_{2n}\| \geq L_n(\theta_0) - \lambda_n\|\tilde{\theta}_{2n}\|^{-\mu}\|\theta_{20}\|.$$

When $\theta_{20} = 0$, if $\tilde{\theta}_{2n} \neq 0$,

$$Q_n(\hat{\theta}_n) = L_n(\hat{\theta}_n) - \lambda_n\|\tilde{\theta}_{2n}\|^{-\mu}\|\hat{\theta}_{2n}\| \geq L_n(\theta_0) - \lambda_n\|\tilde{\theta}_{2n}\|^{-\mu}\|\theta_{20}\| = L_n(\theta_0);$$

if $\tilde{\theta}_{2n} = 0$, $Q_n(\hat{\theta}_n) = L_n(\hat{\theta}_{1n}, 0) \geq L_n(\theta_0)$. Thus, w.p.a.1., for any $\delta > 0$,

$$Q_n(\hat{\theta}_n) > L_n(\theta_0) - \frac{\delta}{3}.$$

By Lemma 2.4 in Newey and McFadden (1994), $\sup_{\theta\in\Theta}|L_n(\theta) - \mathrm{E}\, l_i(\theta)| = o_p(1)$ under Assumption 1. Hence, w.p.a.1.,

$$\mathrm{E}\, l_i(\hat{\theta}_n) \geq L_n(\hat{\theta}_n) - \frac{\delta}{3} \geq Q_n(\hat{\theta}_n) - \frac{\delta}{3} > L_n(\theta_0) - \frac{2\delta}{3} > \mathrm{E}\, l_i(\theta_0) - \delta.$$

Let $\mathcal{N}$ be any relative open subset of Θ containing θ_0. As $\Theta \cap \mathcal{N}^c$ is compact and $\mathrm{E}\, l_i(\theta)$ is uniquely maximized at θ_0, for some $\theta^* \in \Theta \cap \mathcal{N}^c$, $\sup_{\theta\in\Theta\cap\mathcal{N}^c}\mathrm{E}\, l_i(\theta) = \mathrm{E}\, l_i(\theta^*) < \mathrm{E}\, l_i(\theta_0)$. Therefore, choosing $\delta = \mathrm{E}\, l_i(\theta_0) - \sup_{\theta\in\Theta\cap\mathcal{N}^c}\mathrm{E}\, l_i(\theta)$, it follows that w.p.a.1. $\mathrm{E}\, l_i(\hat{\theta}_n) > \sup_{\theta\in\Theta\cap\mathcal{N}^c}\mathrm{E}\, l_i(\theta)$. Thus, the consistency of $\hat{\theta}_n$ follows. □

[11] Note that we cannot use the delta method because $r^{1/3}$ is not differentiable at $r = 0$.

Proof of Proposition 2. Let $\alpha_n = n^{-1/2} + \lambda_n$. As in Fan and Li (2001), we show that for any given $\epsilon > 0$, there exists a large enough constant C such that

$$\mathrm{P}\{\sup_{\|u\|=C} Q_n(\theta_0 + \alpha_n u) < Q_n(\theta_0)\} \geq 1 - \epsilon. \tag{A35}$$

We consider the two cases $\theta_{20} \neq 0$ and $\theta_{20} = 0$ separately.

(i) $\theta_{20} \neq 0$. Note that Taylor's theorem still holds when some parameters are on the boundary (Andrews 1999, Theorem 6) as the parameter space is convex. Then by a first order Taylor expansion of u at 0, w.p.a.1.,

$$\begin{aligned}
&Q_n(\theta_0 + \alpha_n u) - Q_n(\theta_0) \\
&= \alpha_n \frac{\partial L_n(\theta_0)}{\partial \theta'} u + \frac{1}{2}\alpha_n^2 u' \frac{\partial^2 L_n(\theta_0 + \alpha_n \bar{u})}{\partial \theta \partial \theta'} u - \alpha_n \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\theta_{20}\|^{-1} \theta_{20}' u_2 \\
&\quad - \frac{1}{2}\alpha_n^2 \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} u_2' [-\|\theta_{20} + \alpha_n \bar{u}_2\|^{-3} (\theta_{20} + \alpha_n \bar{u}_2)(\theta_{20} + \alpha_n \bar{u}_2)' + \|\theta_{20} + \alpha_n \bar{u}_2\|^{-1} I_p] u_2,
\end{aligned}$$

where u_2 is the subvector of u that consists of the last p elements of u, and $\bar{u}$ lies between u and 0. The first term on the r.h.s. excluding u has the order $O_p(n^{-1/2}\alpha_n) = O_p(\alpha_n^2)$. As $\frac{\partial^2 L_n(\theta_0 + \alpha_n \bar{u})}{\partial \theta \partial \theta'} = \mathrm{E}\frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta'} + o_p(1)$, the second term on the r.h.s. excluding u' and u has the order $O_p(\alpha_n^2)$. The third term on the r.h.s. excluding u_2 has the order $O_p(\lambda_n \alpha_n) = O_p(\alpha_n^2)$, since $\tilde{\theta}_{2n} = \theta_{20} + o_p(1)$ and $\theta_{20} \neq 0$. By the Cauchy-Schwarz inequality, the fourth term on the r.h.s. is bounded by $\alpha_n^2 \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} u_2' u_2 \|\theta_{20} + \alpha_n \bar{u}_2\|^{-1} = O_p(\lambda_n \alpha_n^2) = o_p(\alpha_n^2)$. Since $\mathrm{E}\frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta'}$ is negative definite, for a sufficiently large C, the second term dominates other terms. Thus, (A35) holds.

(ii) $\theta_{20} = 0$. If $\tilde{\theta}_{2n} = 0$, then $Q_n(\theta) = L_n(\theta_1, 0)$ and the PMLE becomes the restricted MLE with $\theta_2 = 0$ imposed. Thus, $\hat{\theta}_n = O_p(n^{-1/2})$. If $\tilde{\theta}_{2n} \neq 0$, then $Q_n(\theta) = L_n(\theta) - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\theta_2\|$ and

$$\begin{aligned}
Q_n(\theta_0 + \alpha_n u) - Q_n(\theta_0) &= L_n(\theta_0 + \alpha_n u) - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\alpha_n u_2\| - L_n(\theta_0) \\
&\leq L_n(\theta_0 + \alpha_n u) - L_n(\theta_0).
\end{aligned}$$

Expanding $L_n(\theta_0 + \alpha_n u) - L_n(\theta_0)$ by Taylor's theorem as in (i), we see that (A35) holds.

Equation (A35) implies that there exists a local maximum in the ball $\{\theta_0 + \alpha_n u : \|u\| \leq C\}$ with probability at least $1 - \epsilon$. Furthermore, for given $\epsilon > 0$, because $\hat{\theta}_n$ is a consistent estimator of θ_0 by Proposition 1, there exists a small ball with radius $\delta > 0$, such that $P(\|\hat{\theta}_n - \theta_0\| \leq \delta) \geq 1 - \epsilon$. So one may choose C such that the small ball is a subset of $\{\theta_0 + \alpha_n u : \|u\| \leq C\}$ and (A35) holds. Because $Q_n(\hat{\theta}_n) \geq Q_n(\theta_0)$, this implies that $\hat{\theta}_n \in \{\theta_0 + \alpha_n u : \|u\| \leq C\}$. Then the result in the proposition holds. □

Proof of Proposition 3. From the construction of $Q_n(\theta)$ in (1), if the initial $\tilde{\theta}_{2n} = 0$, $\hat{\theta}_{2n}$ is set to zero. So it is sufficient to consider $\tilde{\theta}_{2n} \neq 0$. If $\hat{\theta}_{2n} \neq 0$, we have the first order condition

$$\frac{\partial L_n(\hat{\theta}_n)}{\partial \theta_2} - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \hat{\theta}_{2n} \|\hat{\theta}_{2n}\|^{-1} = 0. \tag{A36}$$

By a first order Taylor expansion, $\frac{\partial L_n(\hat{\theta}_n)}{\partial \theta_2} = \frac{\partial L_n(\theta_0)}{\partial \theta_2} + \frac{\partial^2 L_n(\check{\theta}_n)}{\partial \theta_2 \partial \theta'}(\hat{\theta}_n - \theta_0)$, where $\check{\theta}_n$ lies between θ_0 and $\hat{\theta}_n$. Let $\mathcal{T}$ be a relative compact neighborhood of θ_0 contained in $\mathcal{S}$. Under Assumption 4, by Lemma 2.4 in Newey and McFadden (1994), $\sup_{\theta \in \mathcal{T}} \|\frac{\partial L_n(\theta)}{\partial \theta_2} - \mathrm{E}\frac{\partial l_i(\theta)}{\partial \theta_2}\| = o_p(1)$, $\sup_{\theta \in \mathcal{T}} \|\frac{\partial^2 L_n(\theta)}{\partial \theta_2 \partial \theta'} - \mathrm{E}\frac{\partial^2 l_i(\theta)}{\partial \theta_2 \partial \theta'}\| = o_p(1)$, and $\mathrm{E}\frac{\partial l_i(\theta)}{\partial \theta_2}$ and $\mathrm{E}\frac{\partial^2 l_i(\theta)}{\partial \theta_2 \partial \theta'}$ are continuous for $\theta \in \mathcal{T}$. For $L_n(\theta)$ on $\mathcal{S}$, Lemma 3.6 in Newey and McFadden (1994) holds and $\mathrm{E}\frac{\partial l(\theta_0)}{\partial \theta} = 0$. Then $\frac{\partial L_n(\theta_0)}{\partial \theta} = O_p(n^{-1/2})$ as its variance has the order $O(n^{-1})$. As $\mathcal{S}$ is compact, $\frac{\partial^2 L_n(\check{\theta}_n)}{\partial \theta_2 \partial \theta'} = O_p(1)$. Thus, $\frac{\partial L_n(\hat{\theta}_n)}{\partial \theta_2} = o_p(1)$. Furthermore,

if the information matrix is nonsingular, by Proposition 2, $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2} + \lambda_n)$ and $\frac{\partial L_n(\hat{\theta}_n)}{\partial \theta_2} = O_p(n^{-1/2} + \lambda_n)$. Since $\hat{\theta}_{2n} \neq 0$, there must be some component $\hat{\theta}_{2n,j}$ of $\hat{\theta}_{2n} = (\hat{\theta}_{2n,1}, \ldots, \hat{\theta}_{2n,p})'$, where p is the length of θ_2, such that $|\hat{\theta}_{2n,j}| = \max\{|\hat{\theta}_{2n,i}| : 1 \leq i \leq p\}$. Then $|\hat{\theta}_{2n,j}|/\|\hat{\theta}_{2n}\| \geq 1/\sqrt{p} > 0$. Under Assumption 5 (i), the first term on the l.h.s. of (A36) has the order $o_p(1)$, but the maximum of the components in absolute value of the second term goes to infinity w.p.a.1., then (A36) cannot hold with a positive probability. Under Assumption 5 (ii), the first term on the l.h.s. of (A36) multiplied by $n^{-1/2}$ has the order $O_p(1)$, but the maximum of the components in absolute value of the second term multiplied by $n^{-1/2}$ goes to infinity w.p.a.1., then (A36) cannot hold with a positive probability either. Hence, $\mathrm{P}(\hat{\theta}_{2n} = 0) \to 1$ as $n \to \infty$.

Since $\lim_{n\to\infty} \mathrm{P}(\hat{\theta}_{2n} = 0) = 1$, w.p.a.1., we have the first order condition $\frac{\partial L_n(\hat{\theta}_{1n}, 0)}{\partial \theta_1} = 0$. By the mean value theorem,

$$0 = \frac{\partial L_n(\theta_0)}{\partial \theta_1} + \frac{\partial^2 L_n(\bar{\theta}_{1n}, 0)}{\partial \theta_1 \partial \theta_1'}(\hat{\theta}_{1n} - \theta_{10}),$$

where $\bar{\theta}_{1n}$ lies between $\hat{\theta}_{1n}$ and θ_{10}. Thus,

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) = \Big(-\frac{\partial^2 L_n(\bar{\theta}_{1n}, 0)}{\partial \theta_1 \partial \theta_1'}\Big)^{-1} \sqrt{n} \frac{\partial L_n(\theta_0)}{\partial \theta_1}.$$

Under Assumption 4, $\frac{\partial^2 L_n(\bar{\theta}_{1n}, 0)}{\partial \theta_1 \partial \theta_1'} = \mathrm{E}\big(\frac{\partial^2 l_i(\theta_0)}{\partial \theta_1 \partial \theta_1'}\big) + o_p(1)$ and the information matrix equality $\mathrm{E}\big(\frac{\partial^2 l_i(\theta_0)}{\partial \theta_1 \partial \theta_1'}\big) = -\mathrm{E}\big(\frac{\partial l_i(\theta_0)}{\partial \theta_1} \frac{\partial l_i(\theta_0)}{\partial \theta_1'}\big)$ holds, thus the result in the proposition follows. □

Proof of Proposition 4. When $\theta_{20} \neq 0$, by Proposition 2, $\hat{\theta}_n = \theta_0 + O_p(n^{-1/2})$ under Assumption 6, and also $\tilde{\theta}_{2n} \neq 0$ w.p.a.1. As $\theta_0 \in \mathrm{int}(\Theta)$, we have the first order condition

$$\frac{\partial L_n(\hat{\theta}_n)}{\partial \theta} - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\hat{\theta}_{2n}\|^{-1} \begin{pmatrix} 0 \\ \hat{\theta}_{2n} \end{pmatrix} = 0.$$

Applying the mean value theorem to the first term on the l.h.s. yields

$$\frac{\partial L_n(\theta_0)}{\partial \theta} + \frac{\partial^2 L_n(\bar{\theta}_n)}{\partial \theta \partial \theta'}(\hat{\theta}_n - \theta_0) - \lambda_n \|\tilde{\theta}_{2n}\|^{-\mu} \|\hat{\theta}_{2n}\|^{-1} \begin{pmatrix} 0 \\ \hat{\theta}_{2n} \end{pmatrix} = 0,$$

where $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0. As in the proof of Proposition 3, $\mathrm{E}\frac{\partial l(\theta_0)}{\partial \theta} = 0$ and $\frac{\partial L_n(\theta_0)}{\partial \theta} = O_p(n^{-1/2})$. The second term on the l.h.s. has the order $O_p(n^{-1/2})$. By Assumption 6, the third term on the l.h.s. has the order $o_p(n^{-1/2})$. Thus,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \Big(-\frac{\partial^2 L_n(\bar{\theta}_n)}{\partial \theta \partial \theta'}\Big)^{-1} \sqrt{n} \frac{\partial L_n(\theta_0)}{\partial \theta} + o_p(1).$$

Since $\frac{\partial^2 L_n(\bar{\theta}_n)}{\partial \theta \partial \theta'} = \mathrm{E}\big(\frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta'}\big) + o_p(1)$ and the information matrix equality $\mathrm{E}\big(\frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta'}\big) = -\mathrm{E}\big(\frac{\partial l_i(\theta_0)}{\partial \theta} \frac{\partial l_i(\theta_0)}{\partial \theta'}\big)$ holds, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has the asymptotic distribution in the proposition. □

Proof of Proposition 5. We consider the following two cases separately: (1) $\theta_{20} \neq 0$, but $\hat{\theta}_{2\lambda} = 0$; (2) $\theta_{20} = 0$, but $\hat{\theta}_{2\lambda} \neq 0$.

Case 1: $\theta_{20} \neq 0$, but $\hat{\theta}_{2\lambda} = 0$. Let $\check{\theta}_n = (\check{\theta}_{1n}', 0)'$ be the restricted MLE with the restriction $\theta_2 = 0$ imposed, where $\check{\theta}_{1n} = \arg\max_{\theta_1 \in \Theta_1} L_n(\theta_1, 0)$. As $\theta_{20} \neq 0$, $\bar{\theta} \equiv \mathrm{plim}_{n\to\infty} \check{\theta}_n \neq \theta_0$. Then $\mathrm{E}\, l(\bar{\theta}) < \mathrm{E}\, l(\theta_0)$. By the setting of Case 1 and the definition of $\check{\theta}_n$, since $\Gamma_n \to 0$ as $n \to \infty$, $H_n(\lambda) = L_n(\hat{\theta}_\lambda) + \Gamma_n \leq L_n(\check{\theta}_n) + \Gamma_n = \mathrm{E}\, l(\check{\theta}_n) + o_p(1) = \mathrm{E}\, l(\bar{\theta}) + o_p(1)$. Furthermore, by Proposition 2, $\hat{\theta}_{\bar{\lambda}_n} = \theta_0 + o_p(1)$. Then w.p.a.1., $H_n(\bar{\lambda}_n) = L_n(\hat{\theta}_{\bar{\lambda}_n}) = \mathrm{E}\, l(\hat{\theta}_{\bar{\lambda}_n}) + o_p(1) = \mathrm{E}\, l(\theta_0) + o_p(1)$. Hence, $\mathrm{P}(\sup_{\{\lambda \in \Lambda : \theta_{20} \neq 0, \text{ but } \hat{\theta}_{2\lambda} = 0\}} H_n(\lambda) < H_n(\bar{\lambda}_n)) \to 1$ as $n \to \infty$.

Case 2: $\theta_{20} = 0$, but $\hat{\theta}_{2\lambda} \neq 0$. As $\hat{\theta}_{2\lambda} \neq 0$, $H_n(\lambda) = L_n(\hat{\theta}_\lambda)$. By the definition of the MLE $\tilde{\theta}_n$, $L_n(\hat{\theta}_\lambda) \leq L_n(\tilde{\theta}_n)$. By Proposition 3, $\mathrm{P}(\hat{\theta}_{2\bar{\lambda}_n} = 0) \to 1$ as $n \to \infty$, and $\hat{\theta}_{1\bar{\lambda}_n} = \theta_{10} + O_p(n^{-1/2})$. Then w.p.a.1., $H_n(\bar{\lambda}_n) = L_n(\hat{\theta}_{1\bar{\lambda}_n}, 0) + \Gamma_n$. By a first order Taylor expansion (Andrews 1999, Theorem 6), w.p.a.1.,

$$\begin{aligned} n^{2s}[H_n(\lambda) - H_n(\bar{\lambda}_n)] &\leq n^{2s}[L_n(\tilde{\theta}_n) - L_n(\theta_0)] - n^{2s}[L_n(\hat{\theta}_{1\bar{\lambda}_n}, 0) - L_n(\theta_0)] - n^{2s}\Gamma_n \\ &= n^{2s}\frac{\partial L_n(\theta_0)}{\partial \theta'}(\tilde{\theta}_n - \theta_0) + \frac{1}{2}n^s(\tilde{\theta}_n - \theta_0)'\frac{\partial^2 L_n(\ddot{\theta}_n)}{\partial\theta\partial\theta'}n^s(\tilde{\theta}_n - \theta_0) \\ &\quad - n^{2s}\frac{\partial L_n(\theta_0)}{\partial \theta_1'}(\hat{\theta}_{1\bar{\lambda}_n} - \theta_{10}) - \frac{1}{2}n^{2s}(\hat{\theta}_{1\bar{\lambda}_n} - \theta_{10})'\frac{\partial^2 L_n(\breve{\theta}_n)}{\partial\theta_1\partial\theta_1'}(\hat{\theta}_{1\bar{\lambda}_n} - \theta_{10}) - n^{2s}\Gamma_n, \end{aligned}$$

where $\ddot{\theta}_n$ lies between θ_0 and $\tilde{\theta}_n$, and $\breve{\theta}_n$ lies between θ_0 and $\hat{\theta}_{\bar{\lambda}_n}$. As in the proof of Proposition 3, $\sup_{\theta\in\mathcal{T}} \| \frac{\partial L_n(\theta)}{\partial\theta} - \mathrm{E}\,\frac{\partial l(\theta)}{\partial\theta} \| = o_p(1)$, $\sup_{\theta\in\mathcal{T}} \| \frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - \mathrm{E}\,\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'} \| = o_p(1)$ and $\frac{\partial L_n(\theta_0)}{\partial\theta} = O_p(n^{-1/2})$. Then the first term on the r.h.s. has the order $O_p(n^{s-1/2}) = O_p(1)$, the second term has the order $O_p(1)$ since $\frac{\partial^2 L_n(\ddot{\theta}_n)}{\partial\theta\partial\theta'} = \mathrm{E}\,\frac{\partial^2 l(\ddot{\theta}_n)}{\partial\theta\partial\theta'} + o_p(1) = \mathrm{E}\,\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'} + o_p(1) = O_p(1)$, the third term has the order $O_p(n^{2s-1}) = O_p(1)$, the fourth term has the order $O_p(n^{2s-1}) = O_p(1)$, and the last term goes to minus infinity as $n \to \infty$. Hence, $\mathrm{P}(\sup_{\{\lambda\in\Lambda:\theta_{20}=0,\text{ but }\hat{\theta}_{2\lambda}\neq 0\}} H_n(\lambda) < H_n(\bar{\lambda}_n)) \to 1$ as $n \to \infty$.

Combining the results in the above two cases, we have the result in the proposition. □

References

Aigner, Dennis, C. A. Knox Lovell, and Peter Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21–37.

Andrews, Donald W. K. 1999. Estimation when a parameter is on a boundary. *Econometrica* 67: 1341–83.

Chen, Jiahua. 1995. Optimal rate of convergence for finite mixture models. *Annals of Statistics* 23: 221–33.

Cox, David R., and Hinkley, David V. 1974. *Theoretical Statistics*. London: Chapman and Hall.

Fan, Jianqing, and Runze Li. 2001. Variable selection via Nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96: 1348–60.

Goldfeld, Stephen M., and Richard E. Quandt. 1975. Estimation in a disequilibrium model and the value of information. *Journal of Econometrics* 3: 325–48.

Jin, Fei, and Lung-Fei Lee. 2017. Irregular N2SLS and LASSO estimation of the matrix exponential spatial specification model. *Journal of Econometrics* forthcoming.

Kiefer, Nicholas M. 1982. *A Remark on the Parameterization of a Model for Heterogeneity*. Working Paper No 278; Department of Economics, Cornell University, Ithaca, NY, USA.

Lee, Lung-Fei. 1993. Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. *Econometric Theory* 9: 413–30.

Lee, Lung-Fei, and Andrew Chesher. 1986. Specification testing when score test statistics are identically zero. *Journal of Econometrics* 31: 121–49.

Newey, Whitney K., and Daniel McFadden. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*. Edited by James J. Heckman and Edward E. Leamer. Amsterdam: Elsevier, chapter 36, vol. 4, pp. 2111–245.

Quandt, Richard E. 1978. Tests of the Equilibrium vs. Disequilibrium Hypotheses. *International Economic Review* 19: 435–52.

Rao, Calyampudi Radhakrishna. 1973. *Linear Statistical Inferene and Its Applications*. New York: John Wiley and Sons.

Rothenberg, Thomas J. 1971. Identification in parametric models. *Eonometrica* 39: 577–91.

Rotnitzky, Andrea, David R.Cox, Matteo Bottai, and James Robins. 2000. Likelihood-based inference with singular information matrix. *Bernoulli* 6: 243–84.

Sargan, John D. 1983. Identification and lack of identification. *Econometrica* 51: 1605–33.

Silvey, Samuel D. 1959. The Lagrangean multiplier test. *Annals of Mathematical Statistics* 30: 389–407.

Wang, Hansheng, and Chelei Leng. 2007. Unified LASSO Estimation by Least Squares Approximation. *Journal of the American Statistical Association* 102: 1039–48.

Wang, Hansheng, and Chelei Leng. 2008. A note on adaptive group lasso. *Computational Statistics and Data Analysis* 52: 5277–86.

Wang, Hansheng, Bo Li, and Chelei Leng. 2009. Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 71: 671–83.

Wang,Hansheng, Runze Li, and Chih-Ling Tsai. 2007. Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method. *Biometrika* 94: 553–68.

Yuan, Ming, and Yi Lin. 2006. Model selection and estimation in regression with group variables. *Journal of the Royal Statistical Society, Series B* 68: 49–67.

Zhang, Yiyun, Runze Li, and Chih-Ling Tsai. 2010. Regularization Parameter Selections via Generalized Information Criterion. *Journal of the American Statistical Association* 105: 312–23.

Zou, Hui. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101: 1418–29.