A STOCHASTIC FRONTIER MODEL WITH UNSPECIFIED TIME-VARYING FIRM EFFECTS

DANIEL WIKSTRÖM

Abstract. We propose a new way of estimating a stochastic frontier model with time-varying firm effects. By means of nonparametric smoothing of categorical data we allow for completely unspecified time-varying firm effects. Furthermore it is well established that smoothing of cell probabilities of sparse contingency tables improves the finite sample performance compared to the frequency estimator. We are taking advantage of this in a panel data context with time-varying firm effects. By Monte Carlo simulations it is shown that smoothing of the time-varying firm effects works very well compared to the parametric panel data method proposed by Cornwell, Schmidt & Sickles (1990). The proposed method is applied on Indonesian rice farmer data. The most pronounced difference in our analysis relative to previous studies that analyzed this data is considerable larger average estimated efficiency levels. Thus the farmers appear to be less inefficient than in previous studies.

This is a preliminary version. Please do not cite without the author's permission.

Date: November 2, 2009.

Key words and phrases. Technical Efficiency, Nonparametric kernel estimation, time-varying firm effects, Panel data (*JEL Classification:* C1, C14, C33, C33).

We want to acknowledge Jeffrey Racine and Christopher Parameter for their insightful comments and suggestions. We would also like to thank Helena Hansson and Yves Surry for helpful and inspiring discussions. Any errors are, of course, our own.

1. Introduction

It is common to estimate technological efficiency scores based on stochastic frontier models designed for panel data with large N (number of firms) and small T (number of time periods). Recent examples include Han, Orea & Schmidt (2005), Yao & Shively (2007), Abdulai & Tietje (2007), Ahn, Lee & Schmidt (2007) and Millimet & Collier (2008) among others. However the technical efficiency scores are normally based on varying intercepts which are only consistent in the time dimension. Furthermore, the focus in most studies has been on the N-asymptotics of the estimates of the production function parameters instead on the finite time dimension properties of the firm effects.

The purpose of this paper is to improve the small sample approximations of the technical efficiency scores without imposing restrictive structures on the time-varying firm effects. Unlike existing methods we will leave these effects completely unspecified by means of nonparametric smoothing of categorical variables, albeit in a semiparametric setting.

It is well known that smoothing of sparse cell probabilities of contingency tables improves the finite sample properties compared to the frequency estimator (Simonoff 1983, Burman 1987, Hall & Titterington 1987, Grund 1993, Burman 2004). If the categorical variable has a natural ordering (e.g. bad, neutral, good) or the variable is a discretization of a continuous variable, it makes sense to borrow information from nearby data points. It has also been shown that gains can be made, in finite MSE ("mean squared error") sense, through smoothing of unordered variables by inducing some bias to reduce variance (Brown & Rundell 1985). To smooth the unordered firm effect should give efficiency gains compared to parametric estimation which splits the data for each firm into subsets containing only T observations each.[1] We therefore want to investigate the extent for which this phenomenon applies to panel data estimation of time-varying firm effects.

The main focus in earlier studies is to model time-varying firm effects, in a parametric framework, without imposing restrictive assumptions on the effects. Some kind of restriction is needed to address identification, either by clustering firms into groups or by restricting the time pattern. Cornwell et al. (1990) assume that the firm effects changes quadratically over time. Lee (1993) and Ahn, Lee & Schmidt (2001) model firm effects that change with one unspecified time factor that is common to all firms. Lee (2006) extended the modeling to firm effects that change group-wise over time. Ahn, Lee & Schmidt (2007) use a novel GMM framework to estimate a p-factor model that enables firm effects to change over time by p unspecified time factors.

[1] The parametric approach is a direct analog to the unsmooth and inefficient frequency estimator of contingency tables.

Through nonparametric smoothing of categorical data we can avoid making potentially restrictive assumptions about the bivariate structure of the time-varying firm effects. However naturally this comes with a cost: the estimator of the parameters of the production function will be asymptotically biased under a fixed T assumption.[2] This is because the bias induced by smoothing does not shrink towards zero as N tends to infinity.

The reminder of the paper is organized as follows: Section 2 presents the stochastic frontier model. Section 3 proposes an estimation method for the model in Section 2. Section 4 consists of Monte Carlo simulations. Section 5 provides an empirical example while Section 6 concludes the paper.

2. A stochastic frontier model with time-varying firm inefficiencies

We assume there is an existing optimal technology at time t

$$y_t = f(x_t) \tag{2.1}$$

where y_t is optimal output at time t and x_t is a $k \times 1$ vector of inputs. If no further assumptions are made the observed output for firm i at time t, y_{it}, is either equal or less than $f(x_{it})$. We would like to call the difference inefficiency, however, in real life data there may exist noise (weather, luck, measurement error etc.). We therefore assume the following linear – stochastic – frontier model

$$y_{it} = \delta_t + x_{it}'\beta - u_{it} + \xi_{it} \equiv x_{it}'\beta + c_{it} + \xi_{it}; \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{2.2}$$

where x_{it} is the $k \times 1$ covariates vector, β is a $k \times 1$ coefficient vector and $c_{it} = \delta_t - u_{it}$ is the time-varying firm effects where δ_t is the frontier intercept at time t and u_{it} (> 0) is the measure of technical inefficiency of firm i at time t. ξ_{it} is an error term, for which strict exogeneity is assumed, $E(\xi_{it}|x_{is}, c_{is}) = 0$, for all $s, t \in T$. [3]

3. Estimation of the stochastic frontier model

The model (2.2) is no more than a special case of the commonly used partially linear model. Let us derive an infeasible estimator by first taking the expectation of (2.2)

[2]This is a conjecture based on evidence from Monte Carlo simulations.

[3]Henderson & Simar (2005) have also modeled unspecified time-varying effects but with a fully nonparametric specification of the production function. This is more general but inefficient if the number of covariates is large due to the "curse of dimensionality". They do not make any Monte Carlo simulations to probe the finite sample properties of their estimator. They also argue that their estimator is consistent under Theorem 2.1 in Li & Racine (2003). However as we can judge this theorem does not subsume discrete covariates. Although it could be extended to the case of mixed data as in Hall et al. (2007), we do not see how it applies to the case where the number of discrete cells (i) increases at the same rate as the number of observations (N).

conditioning on c_{it}

$$E(y_{it}|c_{it}) = E(x_{it}|c_{it})'\beta + c_{it} \tag{3.1}$$

then subtract (3.1) from (2.2) gives

$$y_{it} - E(y_{it}|c_{it}) = (x_{it} - E(x_{it}|c_{it}))'\,\beta + \xi_{it}. \tag{3.2}$$

The least squares estimator of β is under the standard assumptions $\sqrt{N}$-consistent and asymptotically normally distributed. To obtain a feasible estimator the unknown functions, $E(x_{it}|c_{it})$ and $E(y_{it}|c_{it})$, have to be estimated. Robinson (1988) showed that by nonparametric regression under some regularity conditions these two expectations could be estimated such that in turn the feasible estimator of β retains many of the properties of the infeasible estimator. However this is for the case of continuous covariates of $g(\cdot)$. Robinson did not rule out the possibility of obtaining $\sqrt{N}$-consistency for the case of discrete covariates with finite support. But this is not fulfilled in our case since i is increasing with N. For the large N and fixed T case $E(y_{it}|c_{it})$ and $E(x_{it}|c_{it})$ can never be estimated consistently (this of course applies for the parametric estimators as well). The information about each firm only increases with T. When N increases we get more firms but not more information about the already existing ones. Our conjecture is therefore that the estimated β is asymptotically biased in fixed T settings.[4]

Nevertheless we argue that this bias will have little practical impact on the finite MSE efficiency of the time-varying firm effects which is the primary interest of this paper since the technical efficiency scores are based on these estimates.

Estimates of the time-varying firm effects, $\hat{c}_{it}$, are obtained by nonparametric kernel regression of $\left(y_{it} - x_{it}'\hat{\beta}\right)$ on (i,t). The estimated level of technical inefficiency for firm i at time t is $\hat{u}_{it} = \hat{\delta}_t - \hat{c}_{it}$ where $\hat{\delta}_t = max_j \;\; \hat{c}_{jt}$.

For all conditional expectations, $E(y_{it} - x_{it}'\hat{\beta}|c_{it})$, $E(y_{it}|c_{it})$ and $E(x_{it}|c_{it})$ the local constant estimator is used

$$\hat{E}(\Upsilon_{js}|c_{it}) = \frac{(NT)^{-1}\sum_{j,s}\Upsilon_{js}L\left((j,s),(i,t),\lambda\right)}{\hat{p}(i,t)} \tag{3.3}$$

where

$$\hat{p}(i,t) = (NT)^{-1}\sum_{j,s} L\left((j,s),(i,t),\lambda\right), \tag{3.4}$$

$$L\left((j,s),(i,t),\lambda\right) = \ell_u \times \ell_o. \tag{3.5}$$

and Υ_{js} stands for either one of the dependent variables $(y_{it} - x_{it}'\hat{\beta})$, y_{it} or x_{it}. In the current setting the product kernel (3.5) consists of a univariate kernel that handles

[4]This is confirmed by the Monte Carlo simulations in Section 4.

unordered data, ℓ_u, and a univariate kernel that handles ordered data, ℓ_o. The unordered kernel is defined as

$$\ell_u\left((j,t),(i,t),\lambda_u\right) = \begin{cases} 1, & j = i \\ \lambda_u, & \text{otherwise.} \end{cases} \tag{3.6}$$

Note that if $\lambda_u = 0$ the kernel is just an indicator function and $\hat{p}(i,t)$ is the unbiased frequency estimator. On the other hand when $\lambda_u = 1$ the kernel is a uniform weighting function and thus the firm effect is smoothed out ($\lambda_u \in [0,1]$). Also note that if $\lambda_u > 0$ we are actually using information across firms.

For the time variable we use a kernel that takes the order into account,

$$\ell_o\left((i,s),(i,t),\lambda_o\right) = \begin{cases} 1, & s = t \\ \lambda_o^{|s-t|}, & s \neq t. \end{cases} \tag{3.7}$$

This ordered kernel has the same properties as the unordered one at the limits of the support. If $\lambda_o = 0$ the kernel is just the indicator function and if $\lambda_o = 1$ we have uniform weights.[5] Additionally note that this kernel uses the order of the variable. It "borrows" information from nearby cells depending on how far away s is from t. This is the main difference to the unordered kernel that uses equal information from all other firms.

The most important implication using these two kernels arises when $\lambda_u, \lambda_o > 0$. In this case information from all NT observations are used to some degree. This will be particularly useful for the estimation of the time-varying firm effects, c_{it}, which normally are based on the T observations of each individual firm. When T is small, which is not uncommon in applied settings, such estimators can be quite inefficient, statistically speaking.

A crucial part of nonparametric analysis is to select an appropriate bandwidth vector. We have considered both least squares cross-validation (LSCV) and AIC$_c$ (corrected Akaike information criterion) (Hurvich et al. 1998). Our Monte Carlo simulations indicate that the latter works better in finite samples. Li & Racine (2004) also compare the LSCV and AIC$_c$ selectors and come to the same conclusion.

4. Monte Carlo Simulations

In this section the finite sample performance of the proposed estimator is compared to the parametric estimator proposed by Cornwell et al. (1990) (hereafter the CSS estimator). We chose the CSS estimator for two reasons. First, it allows for arbitrary dependencies between the regressors and the time-varying firm effects. This type of

[5] These two kernels are adopted from Ouyang et al. (2009) who show in addition that with some probability irrelevant discrete covariates are smoothed out when selecting the bandwidths by least squares cross-validation (LSCV).

model is far more popular among applied researchers than models based on the assumption of no dependencies. Hence we are following accepted practice (see Kumbhakar 1990, for an ML-estimator based on the assumption of no dependencies). Second, the estimator is purely parametric which enables an explicit and correct data-generating process (DGP).[6] We were also considering the more general p-factor estimator proposed by Ahn et al. (2007), however, this method involve critical instrumenting which we would like to avoid both for the complexity this puts on the DGP and for the potential limitation this estimator may encounters due to the scarceness of appropriate instruments in real-life applications.

The Monte Carlo Simulations are divided into three sections. In all three sections the time-pattern of the CSS estimator is a polynomial of degree two. However we alter the time-pattern specifications of the DGP's to compare the performance of the estimator based on the semiparametric partially linear model (henceforth PLM) to the CSS estimator when the latter is correctly specified (the time-pattern of the DGP is a polynomial of degree two), over-specified (the time-pattern of the DGP is linear) and misspecified (the time-pattern of the DGP is a sine curve). The motivation to use a quadratic specification for the time pattern for the CSS estimator is that we have observed the frequent use of this function in applied work. Recent examples include Choi, Stefanou & Stokes (2006), Rodríguez-Álvarez, Tovar & Trujillo (2007), Rungsuriyawiboon & Stefanou (2008) and Weill (2008) among others. The common use of this time-pattern can be derived back to the original paper by Cornwell et al. (1990). The motivation for a quadratic time-pattern made by Cornwell et al. is that it should allow for productivity growth. Many of the following applied papers refer to Cornwell et al. to explain why they have chosen the quadratic specification. We believe that productivity growth can be characterized by a vast bulk of functions and instead of just picking on out of many we prefer to let the data pick one for us.

The general specification for the three different DGP's is

$$y_{kit} = \beta_1 z_{1it} + \beta_2 z_{2it} + c_{kit} + \varepsilon_{it}. \tag{4.1}$$

where $k = 1, 2, 3$ represents the tree DGP's depending on the specification of the time-varying firm effects, c_{kit}. The slope coefficients are set to $\beta_1 = \beta_2 = 1$ and the error term, ε_{it}, is normal distributed with mean zero and unit variance.

The construction of c_{kit} is quite involved since we want to make sure that (4.1) is characterizing a production function and also to make sure that the impact of c_{kit} does not change for different DGP's. To accomplish this we first obtain values from three different time-patterns

[6]The firm effects are of course not explicitly modeled but the time pattern is, in a finite sample this should give efficiency gains compared to the semiparametric model.

$$a_{1it} = 0.6a_{1i} + 0.3a_{2i}t + 0.6a_{3i}t^2, \tag{4.2}$$

$$a_{2it} = 0.6a_{1i} + 0.6a_{2i}t \tag{4.3}$$

and

$$a_{3it} = 0.6a_{1i} + 0.3a_{2i}t + 0.2a_{3i}\sin(2\pi t), \tag{4.4}$$

where $a_{jit} \sim \text{uniform}(0,1)$, $j = 1,2,3$ and $t = 0.1, 0.2, \ldots, 0.1 \times T$ $(T = 5, 10, 20)$. In the next step we standardize all three variables a_{kit}, $k = 1,2,3$ and add the minimum plus 0.1 to each separate variable. Thus the three different time-varying firm effect components, c_{kit}, $j = 1,2,3$, are positive with minimum 0.1 and unit variance. Note that the manipulations of a_{kit} do not change the functional form. Thus c_{1it} is upwards-sloping quadratic, c_{2it} is upwards-sloping linear and c_{3it} is a upwards-sloping sine curve. These specifications are all examples of time-patterns that could characterize productivity growth.

The regressors z_{jit}, $j = 1,2$, are computed by first standardizing

$$x_{1it} = N(0,1) + 1.6c_{it}, \tag{4.5}$$

$$x_{2it} = N(0,1) + 0.6c_{it}, \tag{4.6}$$

and second, to characterize inputs, we make these values positive (and nonzero) by adding the minimum of each variable plus 0.1. Thus the final variables z_{jit}, $j = 1,2$ ranges from 0.1 and upwards with unit variance. Note that correlation is induced between the time-varying firm effects and the regressors ($\text{corr}(c_{it}, z_{1it}) \approx 0.85$ and $\text{corr}(c_{it}, z_{2it}) \approx 0.5$). Finally note that since all variables, z_{1it}, z_{2it} and c_{kit} are positive the conditional mean of the model (4.1) is positive. For a production function the conditional mean of course should be positive otherwise we could have negative output.

For all three DGP's 100 Monte Carlo simulations of all six combinations $N = 50, 100, 200$ and $T = 5, 10, 20$ are conducted.

4.1. **Correctly specified time-pattern.** In this subsection the results from the Monte Carlo simulations based on the DGP with a quadratic time-pattern is presented (i.e. the CSS estimator is based on a correct specification).[7]

Table 1 provides some results on how well the estimator for β_1 of the partially linear model perform.[8] Table 2 shows the counterparts for the CSS model.

[7]Boxplots of the time-varying firm effects for the three DGP's are provided in Appendix 3.

[8]The measures are bias, variance and mean square error.

N	T	Bias	Var.	MSE
50	5	0.162	0.018	0.044
100	5	0.196	0.009	0.048
200	5	0.176	0.004	0.035
50	10	0.132	0.007	0.025
100	10	0.126	0.004	0.020
200	10	0.127	0.002	0.018
50	20	0.056	0.004	0.007
100	20	0.057	0.002	0.005
200	20	0.062	0.001	0.005

Table 1. Summary statistics of Monte Carlo simulations for β_1 of the PLM model (CSS correctly specified)

N	T	Bias	Var.	MSE
50	5	-0.006	0.044	0.044
100	5	0.042	0.016	0.018
200	5	0.006	0.008	0.008
50	10	0.006	0.011	0.011
100	10	0.003	0.005	0.005
200	10	-0.002	0.003	0.003
50	20	-0.004	0.004	0.004
100	20	-0.003	0.002	0.002
200	20	0.004	0.001	0.001

Table 2. Summary statistics of Monte Carlo simulations for β_1 of the correctly specified CSS model

Note the overall features of these two Tables: the bias of the PLM estimates do not decrease with N (but with T), the CSS estimates seem as unbiased as expected and the variance of the PLM and the CSS estimates decreases with both N and T. Judging from Table 1 and Table 2 the PLM slope estimates are less precise than the CSS estimates measured by MSE. The simulations for β_2 (placed in the appendix) show a similar pattern although the PLM estimates are comparable and a bit better when $T = 5$.

The general conclusion is that the CSS method is slightly better when estimating the slope coefficients although the PLM estimates are comparable in a few cases especially when $T = 5$. This result was expected since the CSS estimator is unbiased and uses all NT observations to reduce variance.

Table 3 contains the results for the estimates of the time-varying firm effects for each model, c_{it}.[9] The P-values presented in the two tables are conducted from two-sided Wilcoxon rank sum tests where the null hypothesis is that the location shift parameter equals zero, i.e. that the two methods works equally well.[10]

The PLM estimates of the time-varying firm effects are significantly closer to the true effects (c_{bit}) compared to the CSS estimates when T equals 5 or 10, and when $T = 20$ the results are indistinguishable. This example reveals the merits of smoothing. Despite

[9]The values presented are obtained as the median of $\sum_{i,t} \frac{(\hat{c}_{bit}-c_{bit})^2}{NT}$, $b = 1, \ldots, 100$, where c_{bit} is the true time-varying firm effect for firm i at time t while b indicate the Monte Carlo replication.

[10]The hypothesis tests are based on the two distributions: $\sum_{i,t} \frac{(\hat{c}^{PLM}_{bit}-c_{bit})^2}{NT}$ and $\sum_{i,t} \frac{(\hat{c}^{CSS}_{bit}-c_{bit})^2}{NT}$, $b = 1, \ldots, 100$.

N	T	PLM	CSS	Ratio	P-value
50	5	0.484	0.796	0.608	0.000
100	5	0.469	0.717	0.654	0.000
200	5	0.416	0.681	0.610	0.000
50	10	0.274	0.350	0.783	0.000
100	10	0.262	0.337	0.776	0.000
200	10	0.248	0.318	0.782	0.000
50	20	0.176	0.175	1.009	0.290
100	20	0.166	0.166	1.000	0.850
200	20	0.167	0.165	1.010	0.222

Table 3. Summary statistics of Monte Carlo simulations for the time varying firm effects of each model (CSS correctly specified)

the correct and explicitly modeled time pattern of the CSS model the PLM estimator is better or equally good in finite samples.

The inefficiency levels have also been estimated (see Appendix). The PLM results for these inefficiencies are even stronger, compared to the CSS estimates, than for the time-varying firm effects. The PLM estimates are significantly better for all settings and when $T = 5$ the measure of fit values are at least four times larger for the CSS estimates. The CSS estimator generally estimates the average time-varying effect very well when N is large (it is a consistent estimator of the average when the time-pattern is correctly specified). However the maximum of the time-varying effects, $max_j \, \hat{c}_{jt}$, are generally poorly estimated along with each single time-varying effect (these estimates are only consistent when T is large). And because of this the estimated efficiency levels, $\hat{u}_{it} = max_j \, \hat{c}_{jt} - \hat{c}_{it}$, are poorly estimated.

Thus the estimates of the slopes are generally better for the CSS model while the opposite is true for the time-varying firm effects.

4.2. **Over-specified time-pattern.** The second subsection presents the Monte Carlo results when the CSS model is over-specified. The results presented in Table 4, Table 5 and Table 6 reveal a similar pattern as in the previous subsection, i.e. the estimates of the slope coefficients are generally more precise for the CSS estimator while the opposite is true for the time-varying firm effects.

Unlike the results from the previews subsection the PLM-estimated firm effects are not only significantly better for $T = 5, 10$ but also for $T = 20$.

N	T	Bias	Var.	MSE
50	5	0.153	0.018	0.041
100	5	0.185	0.009	0.044
200	5	0.165	0.004	0.031
50	10	0.090	0.007	0.016
100	10	0.086	0.004	0.012
200	10	0.088	0.002	0.010
50	20	0.028	0.004	0.005
100	20	0.031	0.002	0.003
200	20	0.037	0.001	0.003

TABLE 4. Summary statistics of Monte Carlo simulations for β_1 of the PLM model (CSS over-pecified)

N	T	Bias	Var.	MSE
50	5	-0.006	0.044	0.044
100	5	0.042	0.016	0.018
200	5	0.006	0.008	0.008
50	10	0.006	0.011	0.011
100	10	0.003	0.005	0.005
200	10	-0.002	0.003	0.003
50	20	-0.004	0.004	0.004
100	20	-0.003	0.002	0.002
200	20	0.004	0.001	0.001

TABLE 5. Summary statistics of Monte Carlo simulations for β_1 of the over-specified CSS model

N	T	PLM	CSS	Ratio	P-value
50	5	0.430	0.791	0.544	0.000
100	5	0.430	0.717	0.600	0.000
200	5	0.392	0.681	0.576	0.000
50	10	0.234	0.348	0.672	0.000
100	10	0.220	0.342	0.643	0.000
200	10	0.210	0.318	0.660	0.000
50	20	0.157	0.176	0.888	0.046
100	20	0.150	0.166	0.903	0.000
200	20	0.149	0.166	0.900	0.000

TABLE 6. Summary statistics of Monte Carlo simulations for the time varying firm effects of each model (CSS over-pecified)

4.3. **Misspecified time-pattern.** In the third subsection the CSS model is misspecified when the DGP time-pattern is an upwards-sloping sine curve. Unlike the two other simulations these results favor the PLM estimator as the determinant of the slope coefficients (at least when T is larger than 5). A clear pattern for the CSS estimator is that the estimated slope coefficients as well as the firm effects are getting worse as T

N	T	Bias	Var.	MSE
50	5	0.100	0.017	0.027
100	5	0.129	0.008	0.025
200	5	0.109	0.004	0.016
50	10	0.217	0.008	0.055
100	10	0.210	0.004	0.048
200	10	0.214	0.002	0.048
50	20	0.221	0.005	0.053
100	20	0.174	0.002	0.033
200	20	0.159	0.001	0.026

Table 7. Summary statistics of Monte Carlo simulations for β_1 of the PLM model (CSS misspecified)

N	T	Bias	Var.	MSE
50	5	-0.004	0.044	0.044
100	5	0.043	0.016	0.018
200	5	0.007	0.008	0.008
50	10	0.227	0.009	0.061
100	10	0.214	0.004	0.050
200	10	0.208	0.003	0.047
50	20	0.447	0.005	0.204
100	20	0.285	0.003	0.084
200	20	0.236	0.002	0.057

Table 8. Summary statistics of Monte Carlo simulations for β_1 of the misspecified CSS model

increases. The CSS estimates are converging to biased values and when T is large the bias is worse since the sine curve is more pronounced.

N	T	PLM	CSS	Ratio	P-value
50	5	0.334	0.788	0.424	0.000
100	5	0.333	0.721	0.463	0.000
200	5	0.309	0.681	0.453	0.000
50	10	0.529	1.022	0.518	0.000
100	10	0.528	1.005	0.525	0.000
200	10	0.564	1.082	0.521	0.000
50	20	0.413	3.394	0.122	0.000
100	20	0.338	1.549	0.218	0.000
200	20	0.302	1.121	0.269	0.000

Table 9. Summary statistics of Monte Carlo simulations for the time varying firm effects of each model (CSS misspecified)

Overall, when the CSS model is misspecified the PLM estimator is preferable.

For all three sections the CSS estimates of the inefficiencies generally exaggerates how inefficient a firm really is. This is especially true when $T = 5$. One example is presented in Figure 1. Figure 1 shows the average inefficiencies over time for the CSS estimates and the PLM estimates compared to the true average inefficiencies for the DGP when the CSS specification is correct and $N = 200$ and $T = 5$ (in Appendix Figure 6 and Figure 7 show similar plots for the other two DGP's).

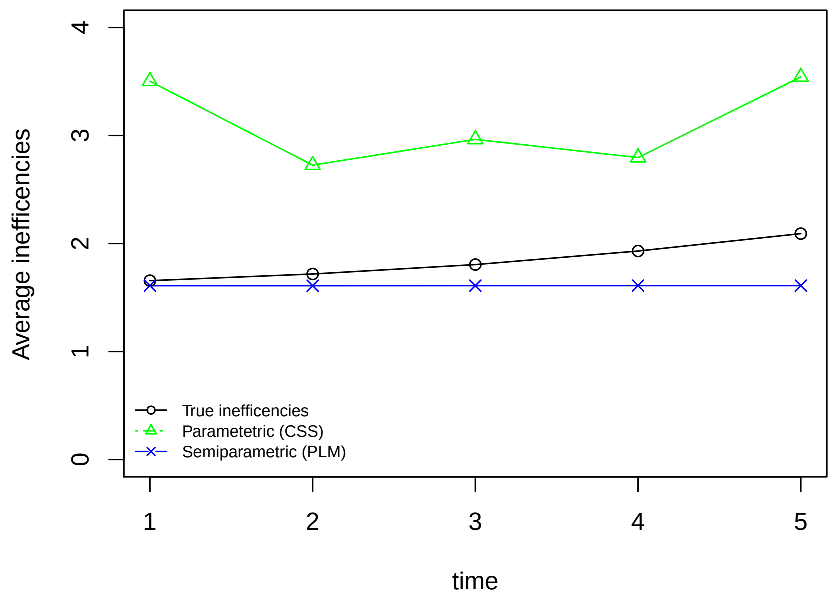

FIGURE 1. Average inefficiency over time when $N = 200$ and $T = 5$ (CSS correctly specified)

For the first time point the CSS average is about twice as large as the true average inefficiency. The CSS-estimated time-varying firm effects are likely to contain a lot of noise when $T = 5$ since the estimator only uses information over T. Noise will cause some extra variability, i.e. the estimated time-varying effects will be more spread out. In turn this will imply that the estimated frontier farm, $max_j\, \hat{c}_{jt}$, will be further away from the other estimates, $\hat{c}_{it}$, and the estimated inefficiencies, $\hat{u}_{it} = max_j\, \hat{c}_{jt} - \hat{c}_{it}$, will be exaggerated.

To summarize all three subsections we conclude that if focus is on estimating technical efficiency scores (or inefficiency levels) based on the time-varying firm effects the PLM estimator is a natural choice. This is true even if the CSS model is correctly specified, in finite samples (small T).

5. Empirical Application: Indonesian rice farmers

In this section Indonesian rice farmer data is analyzed. The Indonesian Ministry of Agriculture surveyed the data from six villages in West Java (Erwidodo 1990). This is a balanced panel of 171 rice farmers over six growing seasons (three wet and three dry).[11] Output is measured in kilograms of rice produced, and inputs are seed (kg), urea (kg), trisodium phosphate (kg), labor (hours) and land (hectares). We assume the commonly used Cobb-Douglas production function. Thus we are using log-transformed inputs and output.[12]

In Table 10 the estimated slope coefficients of the two models are presented. [13] The intra proportions of the PLM estimates are similar to the the intra proportions of the estimated CSS coefficients. The coefficient of *Land* is largest, of *Labor* second largest and so forth. Hence with the CSS estimates as a reference point the PLM estimates seem plausible.

	PLM	CSS
Seed	0.112	0.145
Urea	0.100	0.107
TSP	0.037	0.068
Labor	0.259	0.295
Land	0.442	0.379
In-sample R^2	0.93	0.95
Out-of-sample R^2	0.94	0.80

Table 10. Estimated elasticities

Ahn et al. (2007) concluded that all of the six parametric models they estimate roughly indicated constant returns-to-scale (the sums of the estimated elasticities are around one). Judging from the estimated elasticities in Table 10 this seems valid for the partially linear model as well.

Let us now put attention on the estimates of primary interest, i.e. the technical efficiencies. In Table 11 summary statistics of the CSS estimated firm effects are presented. The CSS estimates show a lot of variation. One farm in period 6 is 15 % as efficient as

[11]One piece of information not considered in previous studies that analyzed this data is gap in the survey between the dry season of 1978 and the wet season 1982/1983. For the PLM estimator we will take this gap into account. This is natural since the ordered kernel is designed to use this kind of information.

[12]Ahn et al. (2007) have an excellent comparison of different panel data estimators based on this data set and since we are using the same functional form the results are directly comparable.

[13]The In-sample $R^2 \equiv \text{corr}(y_{it}, \hat{y}_{it})^2$ is computed on the whole sample while the Out-of-sample R^2 is computed by first randomly splitting the sample into one training sample and one evaluation sample. The training sample is used to obtain estimates of the coefficients (and the bandwidths for the semiparametric model) and in the second step the evaluation data is used together with the estimates obtained from the training data to construct fitted values that in turn are used to obtain $R^2 \equiv \text{corr}(y_{it}, \hat{y}_{it})^2$. The training data consist of 145 farmers while the evaluation contains 26 farms ($N = 145 + 26 = 171$).

	t = 1	t = 2	t = 3	t = 4	t = 5	t = 6
Min.	0.22	0.31	0.32	0.38	0.33	0.15
1st Qu.	0.42	0.47	0.47	0.49	0.45	0.30
Median	0.50	0.55	0.55	0.57	0.54	0.39
Mean	0.53	0.56	0.56	0.58	0.56	0.41
3rd Qu.	0.63	0.63	0.62	0.64	0.63	0.49
Max.	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 11. Summary statistics of estimated technical efficiency scores for the CSS specification

the most efficient farm this period. And for the same period 75 % of the farms are less than 50 % as efficient as the frontier farm (for all time periods the third quartile firm is less than 65 % as efficient). The overall view is that the CSS-estimates appear quite noisy just as could be expected from estimation on six observations ($T = 6$). This is also supported by the in-sample and out-of-sample R^2-values (see Table 10). The out-of-sample R^2 for CSS is 0.95 and it drops to 0.8 when the out-of-sample procedure is used instead. Estimates that contain a lot of noise should not perform well in out-of-sample prediction.

	t = 1	t = 2	t = 3	t = 4	t = 5	t = 6
Min.	0.62	0.66	0.69	0.67	0.60	0.60
1st Qu.	0.75	0.76	0.78	0.79	0.71	0.71
Median	0.80	0.79	0.82	0.83	0.76	0.76
Mean	0.80	0.80	0.82	0.83	0.77	0.77
3rd Qu.	0.86	0.84	0.85	0.87	0.82	0.82
Max.	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 12. Summary statistics of estimated technical efficiency scores for the PLM specification

The technical efficiency estimates for the PLM on the other hand are much less variable (Table 12). For period 6 the least efficient farm is 60 % as efficient as the frontier efficient farm compared to the 15 % for the CSS estimates. The least efficient farm in every time period is actually comparable to the 3^{rd} quartile firms of the CSS estimates. The bandwidths selected by AIC_c are 0.012 for the firm index and virtually zero for the time variable. Thus a small amount of information from all the 171 firms is used to estimate the technical efficiencies for each time period. Put differently the Indonesian rice farmers in West Java have something in common that helps us to reduce the variance in this sample.

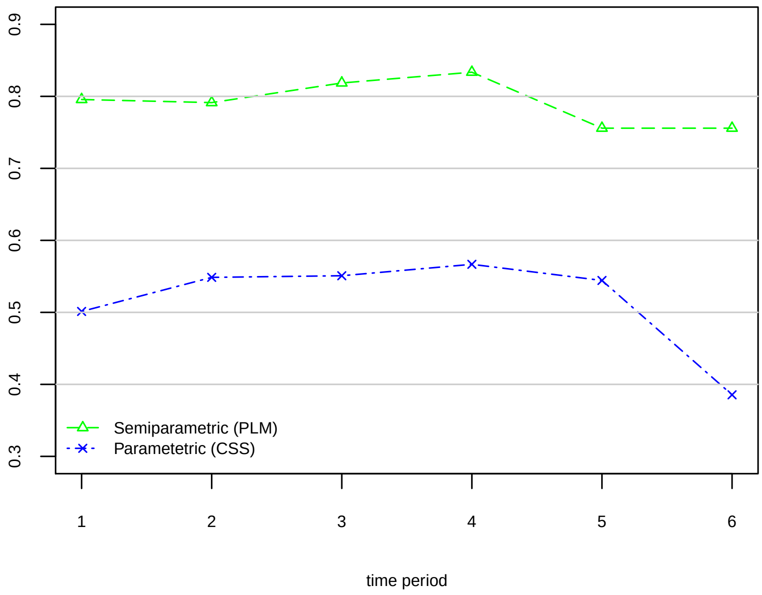

FIGURE 2. Variation of technical efficiency over time

There is also some variation in the time dimension for both models. Figure 2 shows how the average estimated technical efficiencies develop over time. The PLM estimates varies around 0.8 while the CSS estimates varies around 0.55 until period 6 where the curve dips to below 0.4. There is a risk that the CSS method exaggerates the inefficiency of the farmers, just like it did in the Monte Carlo simulations when $T = 5$.

An overall summary of the empirical results could be as follows. Although the estimated elasticities do not differ much between the CSS and the PLM models, the estimated technical efficiencies are very different. The difference is twofold: firstly the CSS estimates appear more variable and secondly the average levels of the PLM estimated efficiencies are higher.

6. Conclusion

In this paper we have proposed a new way to estimate fully unspecified time-varying firm effects by means for existing methodology of smoothing categorical variables.

Monte Carlo simulations indicate that smoothing improves the finite sample accuracy of the estimated time-varying firm effects and the inefficiency levels even if the parametric estimator is based on a correctly specified time pattern. On the other hand the simulations indicate that the smoothed estimator does not improve estimation of the slope coefficients when the time-pattern is correctly or over-specified. In most panel data applications the slope coefficients are the main objective and thus, if one put trust in a special time-pattern one should go ahead and make the parametric estimation. However for stochastic frontier estimation where the time-varying firm effects are of primary interest the proposed estimation method should be considered.

The method is applied to Indonesian rice farmer data. The average level of estimated technical efficiencies is considerably higher compared to previews studies. We argue that this is due to noisy estimates of the time-varying farm effects that are only based on six observations each for the parametric models. The proposed estimator uses some information across farms which make the estimated time-varying farm effects less variable and hopefully less noisy. We find the result quite compelling. To some extent it makes sense that rice farmers in West Java share some common structure. The proposed semi-parametric estimator should be able to smooth out some noise captured by an estimator that does not use this information.

The gains with our estimator are two-fold: no explicit structure has to be induced on the time-varying firm effects and smoothing of these effects may actually accumulate information across firms that even make the estimator superior to estimators based on explicit time-pattern structures.

Usage of traditional panel data analysis to estimate technical efficiency scores is not completely satisfactory since the accuracy depends on T. However as long as there are limited possibilities for obtaining data sets where the time dimension is large, these methods will be used. In the light of this our proposed estimator could contribute to better small sample approximations of technical efficiency scores.

References

Abdulai, A. & Tietje, H. (2007), 'Estimating technical efficiency under unobserved heterogeneity with stochastic frontier models: application to northern german dairy farms', *European Review Of Agricultural Economics* **34**(3), 393–416.

Ahn, S. C., Lee, Y. H. & Schmidt, P. (2001), 'GMM estimation of linear panel data models with time-varying individual effects', *Journal Of Econometrics* **101**(2), 219–255.

Ahn, S. C., Lee, Y. H. & Schmidt, P. (2007), 'Stochastic frontier models with multiple time-varying individual effects', *Journal Of Productivity Analysis* **27**(1), 1–12.

Brown, P. & Rundell, P. (1985), 'Kernel estimates for categorical data', *Technometrics* **27**, 293–299.

Burman, P. (1987), 'Smoothing sparse contingency-tables', *Sankhya-The Indian Journal Of Statistics Series A* **49**, 24–36.

Burman, P. (2004), 'On some testing problems for sparse contingency tables', *J. Multivar. Anal.* **88**(1), 1–18.

Choi, O., Stefanou, S. & Stokes, J. (2006), 'The dynamics of efficiency improving input allocation', *Journal of Productivity Analysis* **25**, 159–171.

Cornwell, C., Schmidt, P. & Sickles, R. C. (1990), 'Production frontiers with cross-sectional and time-series variation in efficiency levels', *Journal Of Econometrics* **46**(1-2), 185–200.

Erwidodo (1990), Panel Data Analysis on Farm-Level Efficiency, Input Demand and Output Supply of Rice Farming in West Java, Indonesia., PhD thesis, Michigan State University.

Grund, B. (1993), 'Kernel estimators for cell probabilities', *Journal Of Multivariate Analysis* **46**(2), 283–308.

Hall, P., Li, Q. & Racine, J. S. (2007), 'Nonparametric estimation of regression functions in the presence of irrelevant regressors', *Review Of Economics And Statistics* **89**(4), 784–789.

Hall, P. & Titterington, D. M. (1987), 'On smoothing sparse multinomial data', *Australian Journal of Statistics* **29**, 19–37.

Han, C., Orea, L. & Schmidt, P. (2005), 'Estimation of a panel data model with parametric temporal variation in individual effects', *Journal Of Econometrics* **126**(2), 241–267.

Henderson, D. J. & Simar, L. (2005), 'A fully nonparametric stochastic frontier model for panel data', *Unpublished manuscript* .

Hurvich, C. M., Simonoff, J. S. & Tsai, C.-L. (1998), 'Smoothing parameter selection in nonparametric regression using an improved akaike information criterion', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **60**(2), 271–293.

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

Lee, Y. (1993), *The measurement of productive efficiency : techniques and applications*, Oxford University Press, chapter A production frontier model with flexible temporal variation in technical inefficiency.

Lee, Y. H. (2006), 'A stochastic production frontier model with group-specific temporal variation in technical efficiency', *European Journal Of Operational Research* **174**(3), 1616–1630.

Li, Q. & Racine, J. (2003), 'Nonparametric estimation of distributions with categorical and continuous data', *Journal Of Multivariate Analysis* **86**(2), 266–292.

Li, Q. & Racine, J. (2004), 'Cross-validated local linear nonparametric regression', *Statistica Sinica* **14**(2), 485–512.

Millimet, D. L. & Collier, T. (2008), 'Efficiency in public schools: Does competition matter?', *Journal Of Econometrics* **145**(1-2), 134–157.

Ouyang, D., Li, Q. & Racine, J. (2009), 'Nonparametric estimation of regression functions with discrete regressors', *Econometric Theory* **25**, 1–42.

Robinson, P. M. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Rodríguez-Álvarez, A., Tovar, B. & Trujillo, L. (2007), 'Firm and time varying technical and allocative efficiency: An application to port cargo handling firms', *International Journal of Production Economics* **109**, 149–161.

Rungsuriyawiboon, S. & Stefanou, S. (2008), 'The dynamics of efficiency and productivity growth in u.s. electric utilities', *Journal of Productivity Analysis* **30**, 177–190.

Simonoff, J. S. (1983), 'A penalty-function approach to smoothing large sparse contingency-tables', *Annals Of Statistics* **11**(1), 208–218.

Weill, L. (2008), 'On the inefficiency of european socialist economies', *Journal of Productivity Analysis* **29**, 79–89.

Yao, R. T. & Shively, G. E. (2007), 'Technical change and productive efficiency: Irrigated rice in the Philippines', *Asian Economic Journal* **21**(2), 155–168.

7. Appendix

N	T	Bias	Var.	MSE
50	5	-0.036	0.007	0.008
100	5	-0.058	0.003	0.006
200	5	-0.052	0.001	0.004
50	10	-0.106	0.003	0.014
100	10	-0.099	0.002	0.011
200	10	-0.099	0.001	0.011
50	20	-0.093	0.001	0.010
100	20	-0.088	0.001	0.008
200	20	-0.101	0.000	0.011

Table 13. Summary statistics of Monte Carlo simulations for β_2 of the PLM model (CSS correctly specified)

N	T	Bias	Var.	MSE
50	5	0.016	0.015	0.015
100	5	-0.002	0.006	0.006
200	5	0.003	0.003	0.003
50	10	-0.004	0.003	0.003
100	10	0.004	0.002	0.002
200	10	0.004	0.001	0.001
50	20	-0.000	0.001	0.001
100	20	0.003	0.001	0.001
200	20	-0.009	0.000	0.000

Table 14. Summary statistics of Monte Carlo simulations for β_2 of the correctly specified CSS model

Swedish University of Agricultural Sciences, Department of Economics, Box 7013, 750 07 Uppsala, Sweden.

E-mail address: Daniel.Wikstrom@ekon.slu.se

N	T	Bias	Var.	MSE
50	5	-0.033	0.007	0.008
100	5	-0.055	0.003	0.006
200	5	-0.047	0.001	0.004
50	10	-0.094	0.003	0.011
100	10	-0.087	0.002	0.009
200	10	-0.087	0.001	0.009
50	20	-0.070	0.001	0.006
100	20	-0.066	0.001	0.005
200	20	-0.078	0.000	0.007

TABLE 15. Summary statistics of Monte Carlo simulations for β_2 of the PLM model (CSS over-specified)

N	T	Bias	Var.	MSE
50	5	0.016	0.015	0.015
100	5	-0.002	0.006	0.006
200	5	0.003	0.003	0.003
50	10	-0.004	0.003	0.003
100	10	0.004	0.002	0.002
200	10	0.004	0.001	0.001
50	20	-0.000	0.001	0.001
100	20	0.003	0.001	0.001
200	20	-0.009	0.000	0.000

TABLE 16. Summary statistics of Monte Carlo simulations for β_2 of the overspecified CSS model

N	T	Bias	Var.	MSE
50	5	-0.022	0.007	0.007
100	5	-0.042	0.002	0.004
200	5	-0.030	0.001	0.002
50	10	-0.033	0.003	0.004
100	10	-0.031	0.002	0.003
200	10	-0.035	0.001	0.002
50	20	-0.104	0.002	0.013
100	20	-0.055	0.001	0.004
200	20	-0.052	0.001	0.003

TABLE 17. Summary statistics of Monte Carlo simulations for β_2 of the PLM model (CSS misspecified)

N	T	Bias	Var.	MSE
50	5	0.016	0.015	0.015
100	5	-0.002	0.006	0.006
200	5	0.003	0.003	0.003
50	10	0.050	0.004	0.006
100	10	0.053	0.002	0.005
200	10	0.052	0.001	0.004
50	20	0.104	0.002	0.012
100	20	0.071	0.001	0.006
200	20	0.049	0.001	0.003

TABLE 18. Summary statistics of Monte Carlo simulations for β_2 of the misspecified CSS model

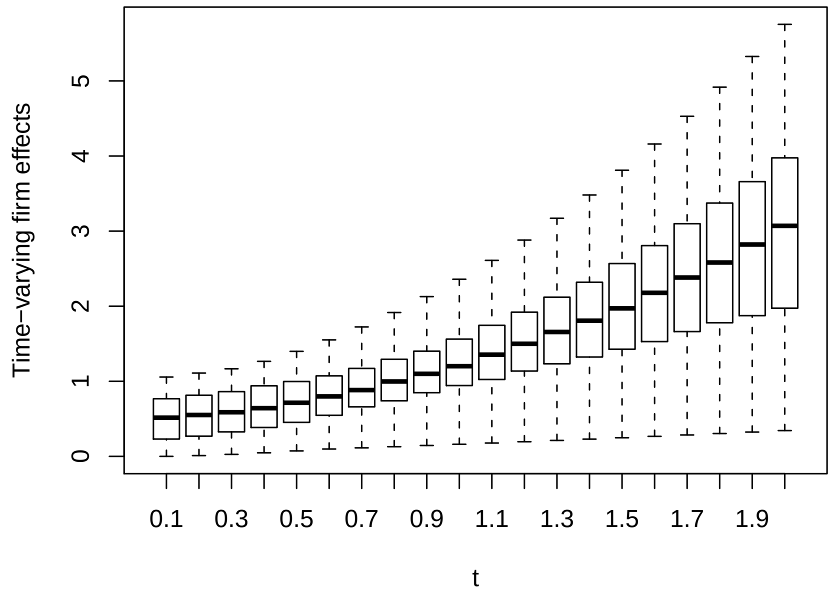

FIGURE 3. The true time-varying firm effects over time for the quadratic time-pattern

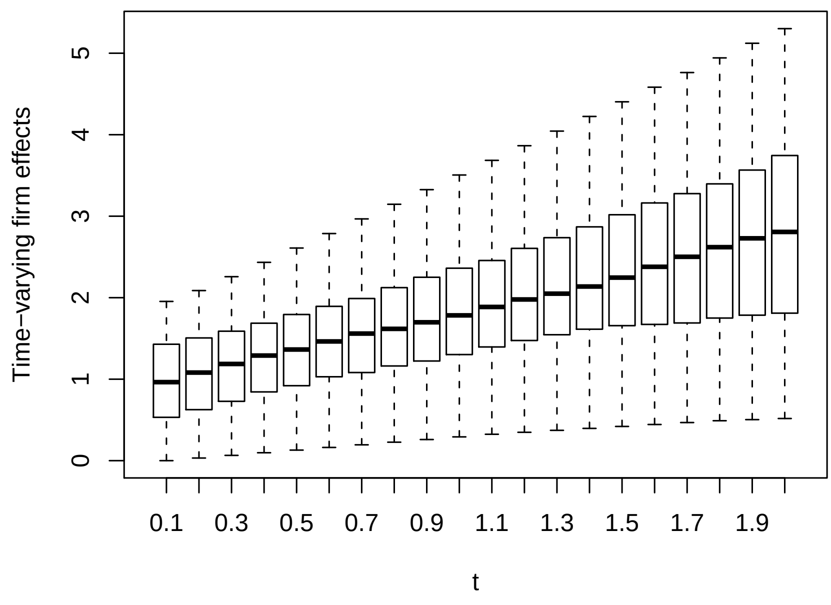

FIGURE 4. The true time-varying firm effects over time for the linear time-pattern

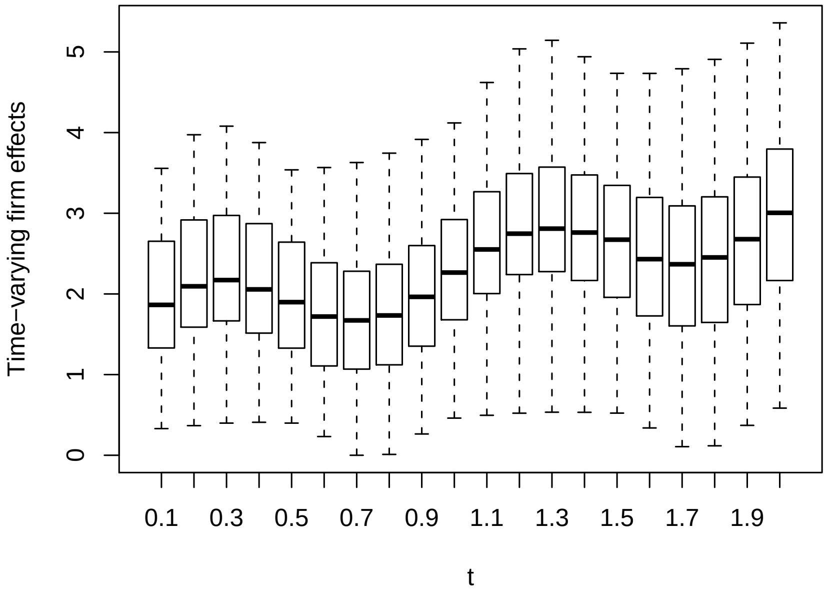

FIGURE 5. The true time-varying firm effects over time for the sine time-pattern

N	T	PLM	CSS	Ratio	P-value
50	5	0.404	1.615	0.250	0.000
100	5	0.412	1.828	0.226	0.000
200	5	0.340	2.346	0.145	0.000
50	10	0.319	0.744	0.429	0.000
100	10	0.348	0.827	0.421	0.000
200	10	0.313	1.017	0.308	0.000
50	20	0.220	0.354	0.620	0.000
100	20	0.229	0.443	0.516	0.000
200	20	0.248	0.509	0.488	0.000

TABLE 19. Measurement of fit (see section 4.1) for the technical inefficiency levels of each model based on the Monte Carlo simulations (CSS correctly specified)

N	T	PLM	CSS	Ratio	P-value
50	5	0.463	1.569	0.295	0.000
100	5	0.438	1.683	0.260	0.000
200	5	0.369	2.251	0.164	0.000
50	10	0.266	0.697	0.382	0.000
100	10	0.252	0.773	0.326	0.000
200	10	0.243	0.949	0.256	0.000
50	20	0.164	0.313	0.522	0.000
100	20	0.184	0.363	0.508	0.000
200	20	0.163	0.456	0.358	0.000

Table 20. Measurement of fit for the technical inefficiency levels of each model based on the Monte Carlo simulations (CSS over-pecified)

N	T	PLM	CSS	Ratio	P-value
50	5	0.361	1.594	0.227	0.000
100	5	0.318	1.757	0.181	0.000
200	5	0.297	2.211	0.134	0.000
50	10	0.429	0.566	0.758	0.000
100	10	0.433	0.590	0.734	0.000
200	10	0.413	0.669	0.617	0.000
50	20	0.405	0.618	0.656	0.000
100	20	0.313	0.385	0.814	0.000
200	20	0.180	0.332	0.542	0.000

Table 21. Measurement of fit for the technical inefficiency levels of each model based on the Monte Carlo simulations (CSS misspecified)

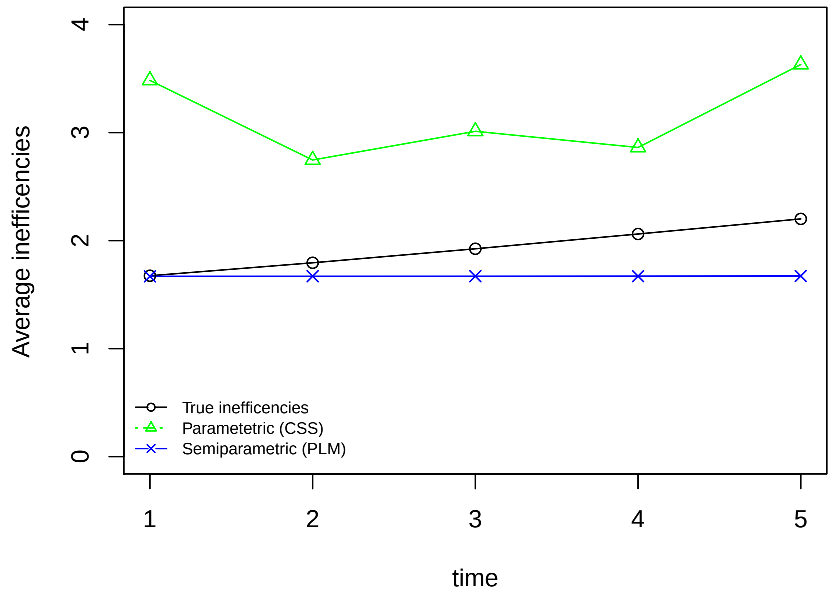

FIGURE 6. Average inefficiency over time when $N = 200$ and $T = 5$ (CSS over-specified)

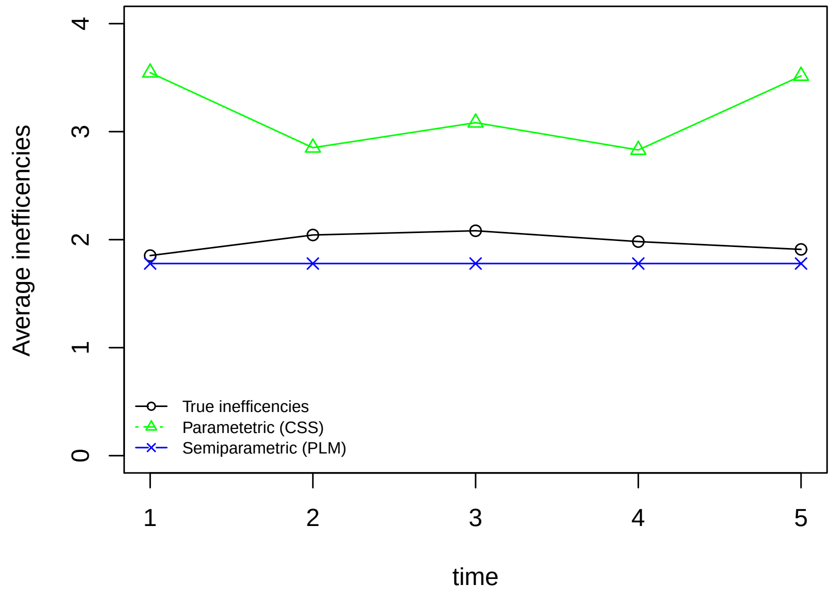

FIGURE 7. Average inefficiency over time when $N = 200$ and $T = 5$ (CSS misspecified)