Syracuse University
SURFACE

Economics Faculty Scholarship | Maxwell School of Citizenship and Public Affairs

10-11-2006

Fixed-Effect Estimation of Technical Efficiency with Time-Invariant Dummies

Qu Feng
Syracuse University

William C. Horrace
Syracuse University, whorrace@maxwell.syr.edu

Follow this and additional works at: http://surface.syr.edu/ecn

Part of the Economics Commons

Recommended Citation
Feng, Qu and Horrace, William C., "Fixed-Effect Estimation of Technical Efficiency with Time-Invariant Dummies" (2006). *Economics Faculty Scholarship*. Paper 11.
http://surface.syr.edu/ecn/11

This Article is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE. It has been accepted for inclusion in Economics Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 95 (2007) 247–252
economics letters
www.elsevier.com/locate/econbase

Fixed-effect estimation of technical efficiency with time-invariant dummies

Qu Feng [a], William C. Horrace [b,*]

[a] Department of Economics, Syracuse University, Syracuse, NY, 13244, USA
[b] Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY, 13244, USA

Received 24 January 2006; received in revised form 25 August 2006; accepted 11 October 2006

Abstract

"Within" estimation of the fixed-effect stochastic frontier model does not identify parameters on time-invariant explanatory variables. If time-invariant variables are important production inputs, then standard efficiency estimates are biased. This note details bias correction, when time-invariant inputs are dummy variables.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Stochastic frontier; Technical efficiency; Bias

JEL classification: C2; D2

1. Introduction

Standard fixed-effect estimation of the stochastic frontier model for panel data was introduced by Schmidt and Sickles (1984). Since then, the specification has been implemented extensively in empirical exercises. Very often, these models have included time-invariant explanatory variables, which the within transformation eliminates. Moreover, the removed time-invariant variables may have been dummy variables. For example, Horrace and Schmidt (2000) estimate a production function for 171 Indonesian rice farms on the island of Java, that includes five dummy variables for six different villages on the island.[1]

* Corresponding author. Tel.: +1 315 443 9061; fax: +1 315 443 1081.
E-mail address: whorrace@maxwell.syr.edu (W.C. Horrace).

[1] The village dummy should not be confused with the village idiot. The distinction is subtle, but should not be overlooked.

0165-1765/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2006.10.013

Their random-effects estimation identifies the village dummy coefficients, but their within estimation does not. This causes their within estimates of farm technical efficiency (unobserved heterogeneity) to be biased. In the words of Greene (2005) the fixed-effect estimator forces "any time invariant cross unit heterogeneity into the same term that is being used to capture the inefficiency."

There are variety of potential solutions to this problem, but all of them effectively move away from the pure simplicity of the fixed-effect model to achieve a solution. Random-effects estimation is possible, but a parametric assumption of the distribution of technical efficiency is required. Non-linear least squares or GMM solutions (e.g., Ahn et al., 2001) add several degrees of computational complexity over within estimation. Here, we detail a simple solution to the problem that remains within the framework of within estimation. Our solution exploits the fact that technical efficiency measures are relative comparisons based on differencing; this differencing can be used to remove the bias caused by unidentified parameters on the time-invariant dummy variables. This is accomplished by making technical efficiency measurements *within particular categories of the dummy variable* (as opposed to across the entire sample), thereby sweeping out the common bias in each dummy variable category. This is how technical efficiency is effectively identified in this note. It should be noted that this method only applies to cases in which the number of time-invariant dummy categories is small relative to the number of firms in the sample, and it does not apply to cases where time-invariant variables are continuous or nearly continuous (countable). We revisit the empirical analysis of Horrace and Schmidt (1996, 2000), and show that although their estimates are biased, their technical efficiency *ranks* are accurate, when compared to the ranks from the techniques detailed herein. We demonstrate theoretically that this is not an artifact of the data.

2. Fixed-effect estimation with dummies

The standard fixed-effect stochastic frontier model with a set of time-invariant dummies is:

$$Y_{it} = \alpha_i + X_{it}\beta + D_i\delta + v_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

where Y and X are a single output and multiple inputs (respectively) of a common production function, parameterized by α, β and δ; $\alpha_i = \eta - u_i$ is a fixed-effect parameter that captures time-invariant technical inefficiency $u_i \geq 0$ (η is the intercept parameter); and v_{it} is an iid zero-mean error term with constant conditional variance, uncorrelated with X. The $D_i = [d_1 \ldots d_{J-1}]$ are $J-1$ time-invariant dummies representing J categories, such that only one element of D_i equals 1 (a complete partition of the data). In our example of Indonesian rice farms, D_i represents 5 dummies for six different villages on the island of Java. The within transformation is:

$$Y_{it} - \overline{Y}_i = (X_{it} - \overline{X}_i)\beta + (v_{it} - \overline{v}_i) \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T.$$

Ordinary least squares provides an unbiased and consistent estimate $\hat{\beta}$. Then inconsistent estimates of α_i are:

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it} = \frac{1}{T}\sum_{t=1}^{T} (Y_{it} - X_{it}\hat{\beta}) \qquad i = 1, \ldots, n.$$

It turns out that: $\underset{T\to\infty}{p\lim}\,\hat{\alpha}_i = \alpha_i + \sum_{j-i}^{J-1} d_j\delta_j$, so if the ith farm is in the jth village, then the probability limit of $\hat{\alpha}_i$ is $\alpha_i+\delta_j$. This is often reported as a consistent estimate of α_i.[2] Therefore, any attempt to estimate relative inefficiency u_i as $\hat{u}_i = \max_{k=1,\ldots,n}\hat{\alpha}_k - \hat{\alpha}_i$ will necessarily be inconsistent, as will the usual normalization of technical efficiency, $\text{TE}_i = \exp\{-\hat{u}_i\} \in [1,0)$. We can, however, consistently estimate relative technical efficiency *within* each of the J categories. Let $N=\{1,\ldots, n\}$, and partition N into, N_j, $j=1,\ldots, J$, so that $N_j \subset N$, and $N = \cup_{j=1}^{J} N_j$. Then a consistent estimate is:

$$\hat{u}_i^* = \max_{k\in N_j}\hat{\alpha}_k - \hat{\alpha}_i, \qquad \text{for all } i\in N_j, \quad \text{and for all } j = 1,\ldots,J.$$

This estimate "differences out" the bias, δ_j. Furthermore, a consistent normalization of relative technical efficiency is:

$$\text{TE}_i^* = \exp\{-\hat{u}_i^*\} \qquad \text{for all } i\in N_j, \quad \text{and for all } j = 1,\ldots,J.$$

Therefore, relative technical efficiency across *all* farms is not identified. However, it is identified, if we simply compare farms within each village.

Empirical implementations typically report the technical efficiency scores, TE_i $1=1,\ldots, N$, in a relative rank statistic, with a single efficient firm having an efficiency score equal to unity. Analogously, we report J rank statistics on TE_i^* $i\in N_j$, for all $j=1,\ldots, J$, so the technique produces J firms having efficiency score equal to unity. It turns out that the within category ranks of the TE_i $i\in N_j$, are identical to the ranks of TE_i^* $i\in N_j$. To see this, we first note that the exponent operator, $\exp\{.\}$, is monotonic in its argument, so we need only compare the ranks of the arguments, $\hat{u}_i$ and $\hat{u}_i^*$. Now, consider some farm $s\in N_j$, then

$$\hat{u}_i - \hat{u}_s = \max_{k=1,\ldots,n}\hat{\alpha}_k - \hat{\alpha}_i - \left(\max_{k=1,\ldots,n}\hat{\alpha}_k - \hat{\alpha}_s\right) = \hat{\alpha}_s - \hat{\alpha}_i,$$

and

$$\hat{u}_i^* - \hat{u}_s^* = \max_{k\in N_j}\hat{\alpha}_k - \hat{\alpha}_i - \left(\max_{k\in N_j}\hat{\alpha}_k - \hat{\alpha}_s\right) = \hat{\alpha}_s - \hat{\alpha}_i.$$

The equivalence of these expressions implies that the rank of $s\in N_j$, relative to $i\in N_j$, remains the same regardless of which technique is employed.

3. Indonesian rice farms

This particular data set of 171 Indonesian rice farms over six growing seasons has been analyzed a number of times, starting with Erwidodo (1990) and, most recently with Ahn et al. (2005). See Horrace

[2] While empirical implementations may not explicitly admit to dropping relevant time-invariant variables, we suspect that time-invariant variables, that would have otherwise been included in a random-effects model, may have simply been ignored by the analyst. This is, of course, not unreasonable, given the limitations of the model. However, it underscores that fact that the usual estimates from fixed-effects models may be biased in most exercises.

Table 1
Ranked technical efficiency for Indonesian rice farms, villages 1, 2 and 3

Village 1 TE_i	Village 1 TE_i^*	Village 2 TE_i	Village 2 TE_i^*	Village 3 TE_i	Village 3 TE_i^*
0.8861	1.0000	0.6757	1.0000	0.7102	1.0000
0.7344	0.8289	0.6454	0.9551	0.6983	0.9833
0.7024	0.7928	0.6355	0.9405	0.6645	0.9356
0.6896	0.7783	0.6334	0.9373	0.6635	0.9342
0.6587	0.7434	0.6245	0.9243	0.6630	0.9335
0.6288	0.7097	0.6164	0.9122	0.6406	0.9020
0.5937	0.6700	0.6148	0.9098	0.6357	0.8952
0.5833	0.6583	0.6121	0.9058	0.6344	0.8932
0.5762	0.6503	0.5688	0.8418	0.6185	0.8709
0.5581	0.6298	0.5669	0.8390	0.5992	0.8437
0.5562	0.6277	0.5535	0.8191	0.5861	0.8253
0.5540	0.6252	0.5459	0.8078	0.5847	0.8233
0.5350	0.6038	0.5452	0.8068	0.5809	0.8179
0.5316	0.6000	0.5441	0.8053	0.5781	0.8140
0.5010	0.5654	0.5362	0.7935	0.5776	0.8133
0.4982	0.5623	0.5244	0.7761	0.5758	0.8108
0.4809	0.5428	0.5163	0.7641	0.5656	0.7965
0.4642	0.5239	0.5039	0.7457	0.5645	0.7949
0.4626	0.5221	0.4925	0.7288	0.5518	0.7769
		0.4863	0.7198	0.5406	0.7612
		0.4771	0.7061	0.5366	0.7555
		0.4678	0.6923	0.5348	0.7530
		0.4668	0.6908	0.5272	0.7423
		0.4473	0.6619	0.5204	0.7327
				0.5202	0.7325
				0.4983	0.7017
				0.4955	0.6978
				0.4910	0.6913
				0.4905	0.6907
				0.4897	0.6895
				0.4738	0.6671
				0.4700	0.6618
				0.4359	0.6138
				0.4272	0.6016
				0.4223	0.5946
				0.4161	0.5859
				0.3655	0.5147

and Schmidt (2000) for a description of the data. The estimation results of the fixed-effect regression of rice production on the explanatory variables (land, labor, fertilizer, variety, pesticide, season, and a village dummy) are contained in Horrace and Schmidt (1996, Table 3). (We do not present the results of the within estimation here.) The only time-invariant regressor is the village dummy, so its coefficient is not identified, creating a bias in the estimate TE_i. Ranked estimation results for biased TE_i and unbiased TE_i^* are presented in Table 1 for villages 1, 2, and 3 and in Table 2 for villages 4, 5, and 6. As expected, the within village ranks are the same for either TE_i or TE_i^*. The tables highlight the magnitude differences

Table 2
Ranked technical efficiency for Indonesian rice farms, villages 4, 5, and 6

Village 4 TE_i	Village 4 TE_i^*	Village 5 TE_i	Village 5 TE_i^*	Village 6 TE_i	Village 6 TE_i^*
0.6460	1.0000	0.9323	1.0000	1.0000	1.0000
0.6394	0.9897	0.6726	0.7215	0.9303	0.9303
0.6234	0.9649	0.6594	0.7073	0.8993	0.8993
0.6089	0.9424	0.6460	0.6929	0.7717	0.7717
0.5849	0.9054	0.6388	0.6853	0.7653	0.7653
0.5831	0.9026	0.6083	0.6525	0.7638	0.7638
0.5805	0.8985	0.6054	0.6493	0.7475	0.7475
0.5728	0.8867	0.5963	0.6396	0.7259	0.7259
0.5629	0.8713	0.5959	0.6392	0.7210	0.7210
0.5616	0.8693	0.5922	0.6352	0.7141	0.7141
0.5518	0.8541	0.5742	0.6160	0.7005	0.7005
0.5449	0.8434	0.5489	0.5888	0.6934	0.6934
0.5425	0.8398	0.5482	0.5881	0.6406	0.6406
0.5251	0.8128	0.5405	0.5797	0.6221	0.6221
0.5236	0.8105	0.5100	0.5470	0.6145	0.6145
0.5221	0.8082	0.5006	0.5370	0.6074	0.6074
0.5191	0.8036	0.4997	0.5360	0.6017	0.6017
0.5159	0.7986	0.4857	0.5210	0.5931	0.5931
0.5065	0.7840	0.4777	0.5124	0.5928	0.5928
0.5053	0.7822	0.4680	0.5020	0.5911	0.5911
0.5025	0.7778	0.4116	0.4415	0.5823	0.5823
0.4929	0.7630	0.3790	0.4065	0.5646	0.5646
0.4800	0.7430			0.5637	0.5637
0.4792	0.7417			0.5631	0.5631
0.4733	0.7326			0.5547	0.5547
0.4715	0.7298			0.5495	0.5495
0.4521	0.6998			0.5313	0.5313
0.4513	0.6985			0.5280	0.5280
0.4504	0.6972			0.5233	0.5233
0.4465	0.6911			0.5160	0.5160
0.4300	0.6655			0.5159	0.5159
0.4208	0.6513			0.4805	0.4805
0.3980	0.6160			0.4543	0.4543
				0.4222	0.4222
				0.4117	0.4117
				0.3837	0.3837

Since the most efficient farm based on TE_i is in the village 6, the biased TE_i is equal to unbiased TE_i^*. This is not the case for the farms in other villages.

between the two estimators of efficiency; these are presumably due to differences in bias. Also as expected, there is a single efficient farm in each village as measured by $TE_i^* = 1.000$, but only one as measured by TE_i, which is in village 6 (Table 2). Notice that in the last two columns of Table 2, the efficiency scores, TE_i and TE_i^*, are identical for village 6. This is because the most efficient farm with $TE_i = 1.000$ happened to be in this village (top of the second to last column). This is not the case for the farms in other villages, where TE_i is always less than unity.

Acknowledgements

We thank Chris Parmeter for encouraging us to publish this note. We blame Ron Oaxaca for the bad pun in footnote 1.

References

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, 219–255.

Ahn, S.C., Lee, Y.H., Schmidt, P., 2005. Panel data models with multiple time-varying individual effects: application to a stochastic frontier production model. Unpublished Manuscript.

Erwidodo, 1990. Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in west Java, Indonesia. Ph.D. Dissertation. Department of Agricultural Economics, Michigan State University.

Greene, W.H., 2005. Fixed and random effects in stochastic frontier models. Journal of Productivity Analysis 23, 7–32.

Horrace, W.C., Schmidt, P., 1996. Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis 7, 257–282.

Horrace, W.C., Schmidt, P., 2000. Multiple comparisons with the best, with economic applications. Journal of Applied Econometrics 15, 1–26.

Schmidt, Sickles, 1984. Production frontiers and panel data. Journal of Business and Economics Statistics 2, 367–374.