HEC : 938
pp.1–22 (col.fig.: NIL)
PROD.TYPE: COM
3B2
ED: CHANDRA ELIZABETH
PAGN: DORTHY SCAN: BINDU

HEALTH ECONOMICS
Health Econ. **13**: 000–000 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:**10.1002/hec.938

Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems

William Greene*
Department of Economics, Stern School of Business, New York University, New York, USA

Summary

The most commonly used approaches to parametric (stochastic frontier) analysis of efficiency in panel data, notably the fixed and random effects models, fail to distinguish between cross individual heterogeneity and inefficiency. This blending of effects is particularly problematic in the World Health Organization's (WHO) panel data set on health care delivery, which is a 191 country, 5-year panel. The wide variation in cultural and economic characteristics of the worldwide sample produces a large amount of unmeasured heterogeneity in the data. This study examines several alternative approaches to stochastic frontier analysis with panel data, and applies some of them to the WHO data. A more general, flexible model and several measured indicators of cross country heterogeneity are added to the analysis done by previous researchers. Results suggest that there is considerable heterogeneity that has masqueraded as inefficiency in other studies using the same data.

JEL classification: C1; C4

Keywords panel data; fixed effects; random effects; random parameters; technical efficiency; stochastic frontier; heterogeneity; health care

Introduction

The *World Health Report* 2000 (*WHR*) [1] is a worldwide assessment of the effectiveness of health care delivery. The study presents a rankings based comparison of the productive efficiency of the health care systems of 191 countries. Predictably, the attention focused on these rankings has been considerably out of proportion to the space this section occupies in the larger report.

The rankings were produced using a form of the 'fixed effects', stochastic frontier methodology proposed by Schmidt and Sickles [2] and Cornwell *et al.* [3] (see [4,5]) (*ETML*). The data analyzed in this econometric study were a 5-year (1993–1997) panel. This section of the WHR has been heavily criticized for numerous reasons related to the overall objectives, the quality and validity of the effectiveness measures, the input data used, and the appropriateness of the methodology (see, e.g. Gravelle *et al.* [6,7] (*GJJS*), Williams [8] and Hollingsworth and Wildman [9] (*HW*)). On January 8, 2001, the authors of the WHO study convened a panel of researchers specifically to discuss the econometric methodology.[a] The focus of the meeting was the use of panel data, such as those in the WHR study, for measurement of

*Correspondence to: Department of Economics, Stern School of Business, New York University, 44 West 4th St., New York, NY 10012, USA. E-mail: wgreene@stern.nyu.edu. URL: www.stern.nyu.edu/~wgreene

Received 30 September 2003
Accepted 1 June 2004

efficiency in health care delivery. This paper reports subsequent research undertaken by the author to study some of the issues raised at that meeting.

One criticism of the fixed effects methodology used is that the model fails to distinguish between cross country heterogeneity unrelated to inefficiency and the inefficiency itself. This ambiguity is likely to be especially problematic in these data, as they are based on 191 sometimes vastly different countries; France, England and Australia appear in the sample on equal terms with Oman, Sri Lanka, Zimbabwe, the Seychelles, Colombia and Bangladesh. We undertook to examine this issue, to reanalyze the WHO data and the methods used in the study, and to propose alternative stochastic frontier-based methods with greater flexibility that will allow the analyst to segregate individual, unmeasured heterogeneity and technical or cost inefficiency.

The paper is organized as follows: The following section reviews the WHO methodology and takes a cursory look at their results. Next, the stochastic frontier model and strategies for modeling panel data are reviewed.[b] The WHO data that were used in this study as well are then described. The several studies of the WHO data [4,5], HW [9], GJJS [6,7] that we examined were based on two output measures, a composite measure of health care delivery (COMP) and disability adjusted life expectancy (DALE) and two inputs, health care expenditure and education levels. In this study, we consider how to use additional country specific covariates including per capita income, income distribution, government effectiveness, and the allocation of health care expenditure between the public and private sectors to account for some of the heterogeneity noted earlier. The empirical results are then presented. We begin with an examination of the production function used, and propose some results that are in broad agreement with others already in the literature regarding the impact of income and the distribution of income on health care outcomes. The second set of results will compare fixed and random effects estimates of technical inefficiency. We find that concerns of ETML [4,5] notwithstanding, for these data, a form of the random effects model appears to be a useful framework for analyzing the WHO data. We then incorporate the country specific heterogeneity in the estimated distribution of technical inefficiency and in the production function itself. Lastly, some conclusions are drawn.

We note at the outset, this paper is intended to document some methodological developments in the estimation of technical inefficiency and to present some specific empirical results to be compared to the WHO study. It is not obvious or certain that this production function approach is the best way to assess the effectiveness of the delivery of health care (see [8] for an example). Moreover, even if so, it is unclear that the actual outcomes measured do indeed result from a ‘production function’ process as described in standard microeconomics, in which health care is the output, spending and education are the inputs, and a process of optimization underlies the observed data. Our main purpose in this study is indicated in the title. The WHO study provides an interesting and well-positioned application in which to examine the econometric issues we present. Although we will be using the ETML model throughout, we take no position on whether this is the optimal model for this purpose. To a large extent, the model used is for illustrative purposes. Future research should provide more evidence on how best to approach the WHO’s objective of assessing the performance of national health systems.

The WHO studies of health care attainment

Evaluation of the effectiveness of health care policies and reforms faces two large obstacles, quantifying goals and objectives so that outcomes can be measured and enumerating inputs in a way that resources can be directed toward them so as to achieve those objectives. The effectiveness study in the WHR is an attempt to measure health care effectiveness in a production function framework.

ETML’s [4,5] preferred methodology was a ‘panel data’, production function estimator based on the framework proposed by Schmidt and Sickles [2] and Cornwell *et al.* [3]. The central feature of the estimator is a fixed effects linear regression model. The production function is denoted as

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} - u_i \qquad (1)$$

where y_{it} is the (log of the) output of the system, $\mathbf{x}_{it}$ is (logs of) the set of inputs, v_{it} is the random component representing stochastic elements as well as any country and time specific heterogeneity, u_i is the inefficiency in the system, and i and t denote country and year, respectively. It is assumed that $u_i > 0$. The equation is rewritten as

$$\begin{aligned} y_{it} &= (\alpha - u_i) + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} \\ &= \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} \end{aligned} \tag{2}$$

Assuming that v_{it} has the familiar stochastic properties of a regression disturbance and is uncorrelated with other components of the model, the parameters can be estimated by least squares, using the 'within', or dummy variable estimator. The country specific constants embody the technical inefficiency. The inefficiencies are estimated in turn by shifting the function upward so that each constant term is measured as a deviation from the benchmark level

$$\hat{u}_i = \max_i(\hat{\alpha}_i) - \hat{\alpha}_i \geq 0 \tag{3}$$

(Note that by this construction, at least one country is measured as 100% efficient.) Technical efficiency is now measured by

$$\mathrm{TE}_{it} = \frac{E[y_{it} \mid \mathbf{x}_{it}, u_i]}{E[y_{it} \mid \mathbf{x}_{it}, u_i = 0]} \tag{4}$$

Overall efficiency is constructed in ETML [4,5] by normalizing this measure to a constructed minimum level of output that would roughly correspond to a system with zero inputs.

Empirical analysis in the two studies used as inputs per capita public and private health care expenditure (HEXP) and average years of education of the population (EDUC). Two measures of health care attainment were analyzed, disability adjusted life expectancy (DALE) and a composite measure of health care delivery (COMP). The production function employed in both cases was

$$\begin{aligned} y_{it} = \alpha_i &+ \beta_1 \log \mathrm{EXP}_{it} + \beta_2 \log \mathrm{EDUC}_{it} \\ &+ \beta_3 \log^2 \mathrm{EDUC}_{it} + v_{it} \end{aligned} \tag{5}$$

where y_{it} is the log of DALE in ETML [4] and the log of COMP in [5]. Table 1 shows some of the estimated results from the two studies. The specific numeric values for DALE, in years, have a ready interpretation, but those for COMP have no clear numeraire. In principle, each TE_i gives the percentage of maximal output above the minimum that is attained in the country. Based on the numeric values of the attainment measures, these would give the proportional potential for improvement.

These data have been reanalyzed by several authors. A variety of strident criticisms were raised in [6], who questioned the methodology and objectives of the entire study as well as the quality

Table 1. Selected WHO estimates of overall efficiency

	DALE[a]		COMP[b]	
Rank	Country	TE[c]	Country	TE[c]
1	Oman	0.992	France	0.994
2	Malta	0.989	Italy	0.991
3	Italy	0.976	San Marino	0.988
4	France	0.974	Andorra	0.982
5	San Marino	0.971	Malta	0.978
25	Costa Rica	0.882	Germany	0.902
50	Uruguay	0.819	Poland	0.793
100	Jordan	0.711	St. Kitts and Nevis	0.643
150	Afghanistan	0.517	Nepal	0.457
187	Malawi	0.196	Nigeria	0.176
188	Botswana	0.183	Dem. Rep. of Congo	0.171
189	Namibia	0.183	Central African Rep.	0.156
190	Zambia	0.112	Myanmar	0.138
191	Zimbabwe	0.080	Sierra Leone	0.000
	United Kingdom (24)	0.883	United Kingdom (18)	0.925
	United States (72)	0.774	United States (37)	0.838

[a] From Evans *et al.* [4, Appendix].
[b] From Evans *et al.* [5, Annex, Table 1].
[c] Normalized to a minimum value at which health 'inputs' are approximately zero.

of the data set and the appropriateness of the outcome measures. Several econometric studies have placed the first (DALE) study under narrower scrutiny. The second (COMP) has, until now, not been similarly examined.

Gravelle *et al*. (GJJS) [6,7] observed that in the sample of 191 countries, 99.8% of the variation in the log of DALE is between, rather than within the groups (countries). The counterparts for the logs of health expenditure (HEXP) and education (EDUC) are 98.9 and 99.8%, respectively. Thus, there is very little actual 'panel data' variation in these data – it differs only slightly from a cross section. GJJS proceeded to fit several models based on the 'between' estimators (country means) and computed alternative adjusted measures of efficiency. They found varying degrees of correlation between their rankings and those in ETML [4,5], ranging from 0.97 down to about 0.39 (see their Table 2). This suggests, as we find below, that the specification can make a considerable amount of difference in the results (GJJS also argued for inclusion of time effects and other expenditure (income minus health expense) in the model. We will return to the model specification in the section technical efficiency estimates from the WHO data).

The stochastic frontier model

The authors of the WHO analyses questioned the distributional assumptions in the stochastic frontier formulation. However, both GJJS and HW found considerable similarity in the results across a number of different formulations that they specified, though theirs sometimes differed markedly from ETML's results. This suggests that specific assumptions about the distribution of efficiencies may be less restrictive than these views suggest. In addition, we submit that the formulations examined previously were narrower than they could have been, and the stochastic frontier model allows the incorporation of cross country heterogeneity in several ways that are likely to bring large benefits in analyzing data as disparate as these.

Pooled data or cross section variants

The essential form of the stochastic production frontier model (see [10,11]) is

$$y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + v_i - u_i \tag{6a}$$

$$v_i \sim N[0, \sigma_v^2] \tag{6b}$$

$$u_i = |U_i| \tag{6c}$$

$$U_i \sim N[0, \sigma_u^2] \tag{6d}$$

A central parameter in the model is the asymmetry parameter, $\lambda = \sigma_u/\sigma_v$; the larger is λ, the greater is the inefficiency component in the model. Parameters are estimated by maximum likelihood, rather than least squares. Estimation of u_i is the central focus of the analysis. With the model

Table 2. Descriptive statistics for variables, 1997 observations

	Non-OECD		OECD		All	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
DALE	54.32	11.73	70.27	3.01	56.83	12.29
COMP	70.30	10.96	89.42	3.97	73.30	12.34
HEXP	249.17	315.11	1498.27	762.01	445.37	616.36
EDUC	5.44	2.38	9.04	1.53	6.00	2.62
GINI	0.399	0.0777	0.299	0.0636	0.383	0.0836
VOICE	−0.195	0.794	1.259	0.534	0.0331	0.926
GEFF	−0.312	0.643	1.166	0.625	−0.0799	0.835
TROPICS	0.596	0.492	0.0333	0.183	0.508	0.501
POPDEN	757.9	2816.3	454.56	1006.7	710.2	2616.5
PUBFIN	56.89	21.14	72.89	14.10	59.40	20.99
GDPC	4449.8	4717.7	18199.07	6978.0	6609.4	7614.8
Sample	161		30		191	

estimated in logarithms, u_i would correspond (to a small degree of approximation) to $1-\mathrm{TE}_i$ given earlier. Individual specific *efficiency* is typically estimated with $\exp(-\hat{u}_i)$. Alternatively, $\hat{u}_i$, itself, provides an estimate of proportional *inefficiency*. With parameter estimates in hand, one can only obtain a direct estimate of $\varepsilon_i = y_i - \alpha - \mathbf{x}_i'\boldsymbol{\beta} = v_i - u_i$. This is translated into an estimate of u_i using Jondrow *et al.*'s (JLMS) [12] formula,

$$E[u_i|\ \varepsilon_i] = \frac{\sigma\lambda}{1+\lambda^2}\left[z_i + \frac{\phi(z_i)}{\Phi(z_i)}\right], \quad z_i = -\varepsilon_i\lambda/\sigma \tag{7}$$

where $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ and $\phi(z)$ and $\Phi(z)$ are the density and CDF of the standard normal distribution, respectively.

The narrow assumption of half normality is a viewed as a significant drawback in this model.[c] HW and others (see [13]) have extended it to a *truncated* normal model by allowing the mean of U_i in (6d) to be nonzero. A major shortcoming of the *truncated* normal model is that the specification of the distribution of u_i still suppresses individual heterogeneity in inefficiency that is allowed, at least in principle, by the fixed effects formulation. We have several observed indicators of this heterogeneity, such as income distribution, per capita GDP, OECD membership, the public share of health care expenditures, etc., and these can be incorporated into the distribution of u_i in ways that the other methods already discussed cannot accommodate. Letting $\mathbf{h}_i$ denote a set of variables that measure the group heterogeneity, we write

$$E[U_i] = \mu_i = \delta_0 + \mathbf{h}_i'\boldsymbol{\delta} \tag{8}$$

The Jondrow *et al.*'s result in (7) is now changed by replacing z_i with $z_i^* = z_i - \mu_i/(\sigma\lambda)$.

Panel data formulations

The fixed effects model. The Schmidt and Sickles [2] formulation

$$y_{it} = (\alpha - u_i) + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}$$

$$= \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} \tag{9a}$$

$$u_i = \max_i(\alpha_i) - \alpha_i \geq 0 \tag{9b}$$

has been used in a number of applications (see [14]). There are two important assumptions built into this model. First, any time invariant heterogeneity will be pushed into α_i and ultimately into the estimate, $\hat{u}_i$. The WHO data span a tremendous range of cultures, economies, and policy settings. This is likely to be a particularly influential aspect of the model for these data. Second, the model assumes that inefficiency is time invariant. For short time intervals, this may be a reasonable assumption. But, five years may be long enough for this to be questionable. HW did find evidence to suggest that this assumption may be inconsistent with the data.

Both of these restrictions can be relaxed by placing country specific constant terms in the stochastic frontier model – we call this a 'true' fixed effects model,

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} - u_{it} \tag{10}$$

where u_{it} has the specifications in (6), (8) for the stochastic frontier model. Superficially, this amounts simply to adding a full set of country dummy variables to the stochastic frontier model. The model is still fit by maximum likelihood, not least squares. This approach is generally avoided, both for the practical difficulty of computing the large number of parameters (the dummy variables cannot be conditioned out of the model) and because of a presumption that the incidental parameters would appear – a persistent bias in estimates of the main parameters in 'true' fixed effects models such as this. Surprisingly, this approach has hardly been used previously[d] in spite of the fact that most of the received panel data applications involved fairly small panels, and there exists almost no actual evidence on the incidental parameters problem in stochastic frontier estimation.[e]

This true fixed effects model places the unmeasured heterogeneity in the production function; with a loglinear model, it produces a neutral shift of the function, specific to each country. One might, instead, have the heterogeneity reside in the inefficiency distribution. This could be accomplished with the modification of (10),

$$\mu_i = \delta_{0i} + \mathbf{h}_i'\boldsymbol{\delta} \tag{11}$$

that is, by placing the country specific dummy variables in the mean of the truncated normal distribution, rather than in the production function. Once again, in a moderate sized sample, this is a minor reformulation of the familiar model. The problems noted in the next paragraph will appear, but how serious these are is an empirical issue, not a foregone conclusion.

The true fixed effects model has the virtues mentioned above. Weighing against it are, first, the incidental parameters problem and second, the possibility that the model is now overspecified. The incidental parameters problem [20] is a persistent bias that arises in nonlinear fixed effects models when the number of periods is small (five is small). It has been widely documented for binary choice models (see, e.g., [21,22]) but not systematically examined in stochastic frontier models. In Greene [Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J Econ* 2004, forthcoming], we found that the biases in coefficient estimates were surprisingly small and did not appear in the patterns predicted by received results for other models, and, moreover, that there appeared to be only minor biases transmitted to the estimates of technical inefficiency. The second problem now is that the model may be overspecified. If there is persistent inefficiency, it is now completely absorbed in the country specific constant term which is also capturing any time invariant heterogeneity. Whereas the earlier fixed effects form would tend to overestimate the inefficiency component, it is possible that this form will underestimate it. Unfortunately, this blending of the two effects is inherent in the modeling problem, and there is no simple solution that will be entirely satisfactory. Ultimately, $\alpha_i + v_{it} - u_{it}$ contains both country specific heterogeneity and inefficiency, and both may have invariant and time varying elements. The intent of this paper is to suggest that accommodating the possibility of time invariant heterogeneity, to the extent possible, is preferable to ignoring it altogether, as in (9).

The random effects model. The random effects model is obtained by assuming that u_i is time invariant and uncorrelated with the included variables in the model

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it} - u_i \tag{12}$$

In the linear regression case analyzed by ETML, the parameters are estimated by two step generalized least squares (see [23, Chapter 11]). On the basis of a Hausman specification test, ETML [4,5] concluded that the random effects model would not be appropriate for their data. Their test was conducted in a model that did not include any of the observed country specific effects, so the result may have been more convincing than appropriate. A significant problem of interpretation would emerge in that test, as the real interest is in whether the inefficiency term could be treated as a random or fixed effect, but the Hausman test could reject the former because of the presence of heterogeneity correlated with the regressors, *but not necessarily related to inefficiency* in the model as such. This complication would be a special feature of the stochastic frontier model. However, even without this difficulty, the linear regression based random effects model would have a significant drawback for present purposes even if it were not rejected; there is no implied estimator of inefficiency in this model, that is, no estimator of TE_i as in the fixed effects case.[f]

Pitt and Lee [24] showed how the time invariant composed error model could be extended to a panel data version of the stochastic frontier model. The direct extension would be of limited usefulness here, first because of the assumption of uncorrelatedness of u_i and $\mathbf{x}_{it}$ and, second, because of the assumption of time invariance of the inefficiency. The first of these can be remedied in the same fashion as suggested earlier. Estimation of the random effects model with heterogeneity in $E[U_i]$, see (8), is straightforward (see [25]). With this extension, the JLMS estimator becomes

$$E[u_i \mid \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T, \mathbf{h}_i] = Z_i + \psi \left[\frac{\phi(Z_i/\psi)}{\Phi(Z_i/\psi)}\right] \tag{13}$$

where $Z_i = \gamma\mu_i - (1-\gamma)\bar{\varepsilon}_i$, $\gamma = 1/(1 + T\lambda^2)$, $\psi^2 = \gamma\sigma_u^2$, and $\bar{\varepsilon}_i = (1/T)\Sigma_i\varepsilon_{it}$.

Battese and Coelli (*BC*) [26,27] have proposed a now widely used modification of the truncation model that can also accommodate some systematic variation in the inefficiency. In their formulation,

$$u_{it} = \eta_t|U_i| \tag{14}$$

where $\eta_t = 1 + \eta_1(t-T) + \eta_2(t-T)^2$ and $U_i \sim N[\mu, \sigma_u^2]$. Various other forms of the function η_t have been proposed, such as $\eta_t = \exp[-\eta(t-T)]$ (see BC [26,27] for discussion). Kumbhakar and Orea [28] suggest a more general form, $\eta_{it} = \exp(\mathbf{g}_{it}'\boldsymbol{\pi})$ where $\mathbf{g}_{it}$ is a set of observable covariates and $\boldsymbol{\pi}$ is another parameter vector to be estimated.[g] This extension subsumes BC's formulation as well as many others. Greene [29] added the heterogeneous truncation form in (11) to this as well. Let $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT})$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$. Estimates of technical inefficiency based on this model follow the same form as those in the

Pitt and Lee model in (13), where Z_i is now

$$Z_i^* = \frac{\sigma_v^2 \mu_i + \sigma_u^2 \boldsymbol{\eta}_i' \boldsymbol{\varepsilon}_i}{\sigma_v^2 + \sigma_u^2 \boldsymbol{\eta}_i' \boldsymbol{\eta}_i} \tag{15}$$

and ψ becomes

$$\psi_i^* = \frac{\sigma_u^2 \sigma_v^2}{\sigma_v^2 + \sigma_u^2 \boldsymbol{\eta}_i' \boldsymbol{\eta}_i} \tag{16}$$

The initial assumptions of homogeneity and time invariance are obviated in this model.

A random effects counterpart to the true fixed effects model in (10) would be a 'true random effects' stochastic frontier model,

$$y_{it} = (\alpha + w_i) + \mathbf{x}_{it}' \boldsymbol{\beta} + v_{it} - u_{it} \tag{17}$$

The time invariant, random constant term embodies the cross country heterogeneity in the production function (not the mean of inefficiency). The one sided inefficiency component now varies freely across time and country. This form of the model overcomes both of the drawbacks noted earlier. Estimation of this model by simulated maximum likelihood methods is discussed in Greene [19,30]. Measured heterogeneity (income distribution, public contribution to health care financing, etc.) could enter this model through two avenues. Simple cross country heterogeneity may affect the location of the frontier, which would be modeled in the form

$$w_i = \mathbf{z}_i' \boldsymbol{\theta} + \omega_i \tag{18}$$

This produces a 'hierarchical' or 'multilevel' model. The heterogeneity may also enter the distribution of u_{it} which can, as before, have mean μ_i or, in principle, even μ_{it} with time variation in the covariates. Country specific estimates of inefficiency are computed using the JLMS [12] formulation. We note that this extension of the model, in which the heterogeneity enters both the production relationship and the mean of the inefficiency, can also be accommodated in both the Pitt and Lee [24] model and the Battese and Coelli [26,27] model, in both cases, as specified earlier.

The World Health Organization WHR data set

The data set used in this analysis were used in ETML [4,5].[h] The full data set is a panel of data observed for 191 member countries of the WHO. The panel data are observed for 5 years, 1993–1997, though 51 of the 191 countries are observed in only one year.[i] The data are more fully described in the WHR and in numerous publications that can be obtained from the WHO website.

Two outcome variables were observed,

DALE disability adjusted life expectancy

COMP composite measure of success in 5 health goals, overall health, health distribution, responsiveness, responsiveness in distribution, fairness in financing. This variable was gathered by survey. See [1] for details and discussion

Natural logs of both outcome variables were used in the analysis to follow (all references to these in regression results are based on logs). GJJS [6,7] expressed some skepticism about using logs for the DALE variable. In the interest of comparability, we have maintained the forms used by the researchers at WHO in this study. The first of the outcome variables is the familiar output measure that was analyzed by HW, GJJS and Williams in addition to ETML [4,5]. The second variable was analyzed in [5], but was not analyzed by the other authors mentioned.

Two variables are modeled as the inputs variables to the production process of health care attainment:

HEXP health expenditure per capita in 1997 ppp$

EDUC average years of schooling

Both input variables entered the production function in log form. The translog model, with squares and the cross product was also considered, but not used by ETML. A restricted form of this model that has appeared in the earlier studies is discussed below. In order to maintain comparability with their study, we have adopted their simpler functional form. Several variables that were observed only for 1997 provide indicators of cross country heterogeneity were:

GINI Gini coefficient, income inequality. This measure ranges from zero (perfect equality) to one (complete inequality). Details may be found at the World Bank

	data website on income and growth, www.worldbank.org/research/growth/dddeisqu.htm. The 'Deininger and Squires Data Set' from which it is derived is also described there
VOICE	World Bank measure of democratization and freedom of political unit. This variable indicates aspects of the political process, civil liberties, political rights and the extent to which citizens participate in the selection of the government. It is measured by surveys
GEFF	measure of government effectiveness, World Bank measure. This refers to the quality of public service provision, bureaucracy, competence of civil servants, independence of civil servants from political pressure and the credibility of government commitment to service policies
TROPICS	dummy variable for tropical location
POPDEN	population density, people per km^2
PUBFIN	percentage of health care paid by the government
GDPC	per capita GDP in 1997 ppp$
OECD	dummy variable for OECD membership[j]

The population density and per capita GDP variables appear in logs in all model results to follow. Further details on the data construction are provided in Appendix A in [31]. Descriptive statistics for the 1997 values of the data used here are given in Table 2.

The original data set contained missing values for approximately 40 countries for the VOICE, GEFF and POPDEN variables. Recent updates of the World Bank website enabled us to complete data for VOICE and GEFF for all but seven countries for VOICE (Cook Islands, San Marino, Niue, Monaco, Nauru, Palau, Tuvalu) and four for GEFF (Cook Islands, Niue, Palau, Laos.) Proxies for these remaining values were obtained by computing fitted values from a regression of the respective values of these variables on all other observed data for the country using the complete data. This represents only 7 and 4 missing values out of 191, so the effect should be minimal in the computed results. Data on population density were missing for 18 countries. The website http://www.geography.about.com/library/cia/blcindex.htm contains year 2000 data on land area, population and many other variables for all of the countries for which our data were missing. The population density was calculated by drawing the year 2000 estimated population back to 1997 using the estimated population growth rate for 3 years, then computing the density in persons per square kilometer. The analysis below is thus based on data for all 191 countries.[k]

Technical inefficiency estimates from the WHO data

The production function

Health expenditure is the most visible input to the health care process, and public health expenditure is a major component of health care policy. There is striking variation in the public share of financing of health expenditure across countries and, as noted by Berger and Messer [BM, 32], across time as well. As can be seen in Table 2, the mean and standard deviation of 59.4 and 21.0 for PUBFIN, respectively, suggest a range of variation of at least 20–100%. Some specific values for the larger economies include 72% for Canada, 25% for China, 92% for the Czech Republic, 77% for France, 78% for Germany, 13% for India, 57% for Italy, 82% for Norway, 78% for Sweden, 97% for the United Kingdom and 44% for the United States. It is unclear, however, how *variation* in the public financing of health care will translate to *variation* in health outcomes. Berger and Messer [32] suggest, for example, that the degree of public financing could lead to either improvement or worsening of the *efficiency* of health care delivery. As they note, this aspect has not received much attention in the empirical literature. The data and models framed for this study will allow us to examine this issue.

A number of researchers have examined the responsiveness of health expenditure to increases in income. Newhouse [33] reported estimates of the income elasticity of health expenditures in the range of 1.15–1.31. Subsequent researchers have examined cross section and pooled time series-cross section data sets with similar results (see [32] for a survey). Our own results based on the WHO data for 1997, shown in Table 3, are consistent, with elasticity estimates of 1.08 for the full sample and 1.23 for the OECD countries. The values in excess of 1.0 suggest that populations value health care as a normal good.

BM recount a series of cross section studies that have found small and insignificant relationships between income levels and health outcomes. Subsequent analyses of income distribution as an alternative explanation have likewise concluded that the distribution of income adds little explanatory power. Our results of linear least squares regressions of our two health outcomes, DALE and COMP (in logs) on health care expenditure, education, the Gini measure of income inequality, and the log of per capita GDP and its square, shown in Table 4, do not agree with these findings. The results suggest that for the poorer, non-OECD countries, there are significant relationships in the expected directions both for per capita GDP and for the income distribution. The results also suggest that the relationships are stronger for

Table 3. Income responsiveness of health care expenditure

	Constant	Income	Education	R^2
Non-OECD	−3.67 (0.251)*	1.02 (0.037)*	0.275 (0.616)*	0.902
OECD	−5.57 (0.747)*	1.23 (0.078)*	0.347 (0.196)*	0.917
All	−4.14 (0.201)*	1.09 (0.031)*	0.268 (0.059)*	0.936

Estimated standard errors in parentheses.
*Indicates significant at the 95% level.

Table 4. Health care outcome regressions[a]

		Constant	Hexp	Educ	Gini	Income	Income^2	R^2
NonOECD	DALE	3.04	0.187	0.017	−0.578	0.098		0.676
		(0.176)*	(0.028)*	(0.033)	(0.148)*	(0.037)*		
		−0.712	0.130	0.033	−0.703	1.046	−0.058	0.715
		(0.832)	(0.029)*	(0.313)	(0.142)*	(0.209)*	(0.013)*	
	COMP	3.56	0.100	0.036	−0.426	0.064		0.769
		(0.101)*	(0.016)*	(0.191)	(0.086)*	(0.0214)*		
		1.94	0.075	0.043	−0.481	0.474	−0.025	0.784
		(0.495)*	(0.017)*	(0.019)*	(0.085)*	(0.124)*	(0.008)	
OECD	DALE	3.82	−0.042	0.046	−0.102	0.0157		0.725
		(1.72)*	(0.271)	(0.244)	(0.078)	(0.0316)		
		−0.261	−0.041	0.036	−0.103	−0.896	−0.046	0.760
		(2.25)	(0.026)	(0.024)	(0.074)	(0.477)	(0.024)	
	COMP	3.91	0.035	0.341	−0.119	0.031		0.846
		(0.137)*	(0.216)	(0.195)	(0.621)	(0.025)		
		−0.507	0.036	0.023	−0.120	0.967	−0.049	0.880
		(1.69)	(0.195)	(0.180)	(0.056)*	(0.358)*	(0.019)*	
All	DALE	3.01	0.197	−0.0078	−0.397	0.105		0.714
		(0.162)*	(0.026)*	(0.029)	(0.125)*	(0.034)*		
		−0.481	−0.128	0.031	−0.625	0.980	−0.054	0.757
		(0.627)	(0.026)*	(0.028)	(0.122)*	(0.156)*	(0.009)*	
	COMP	3.54	0.107	0.024	−0.338	0.068		0.823
		(0.093)*	(0.015)*	(0.017)	(0.072)*	(0.196)		
		1.97	0.076	0.041	−0.440	0.462	−0.024	0.840
		(0.371)*	(0.016)*	(0.0165)*	(0.072)*	(0.092)*	(0.006)*	

Estimated standard errors in parentheses.
[a]DALE, COMP, Expenditure, Education, and Per Capita Income are all in logs. Results are all based on the 1997 data.
*Indicates significant at the 95% level.

non-OECD countries than for OECD countries. The pattern of the quadratic relationships suggests that the improvement in health outcomes provides greater benefits at lower incomes than at higher ones.

The basic production function analyzed here (and in HW) is of the simple form

$$\begin{aligned}\text{Health}_{it} &= f(\text{Education}_{it}, \text{Expenditure}_{it}) + v_{it} - u_{it} \\ &= f(\mathbf{x}_{it}) + v_{it} - u_{it} \qquad (19)\end{aligned}$$

We will augment the production function with two covariates that seem more likely to shift the production function than the mean of inefficiency,

$$\mathbf{z}_i = [\text{Tropics}_i, \text{PopDen}_{i,}] \qquad (20)$$

Finally, in our preferred model, we will allow for time variation (technical change) with the year specific dummy variables,

$$\mathbf{t} = \text{year}_{1993}, \text{year}_{1994}, \text{year}_{1995}, \text{year}^{l}_{1996} \qquad (21)$$

Additional influences on health outcomes that appear in the distribution of the inefficiency term are

$$\mathbf{h}_i = [\text{GEff}_i, \text{Voice}_i, \text{Gini}_i, \text{GDP}_i, \text{PubFin}_i, \text{OECD}_i] \qquad (22)$$

The functional form of the production model remains to be determined. ETML specified a translog model,

$$\begin{aligned}\text{LogHealth} = \alpha &+ \beta_1 \log \text{Exp} + \beta_2 \log \text{Educ} \\ &+ \beta_{11} [\tfrac{1}{2} \log^2 \text{Educ}] \\ &+ \beta_{22} [\tfrac{1}{2} \log^2 \text{Exp}] \\ &+ \beta_{12} [\log \text{Educ})(\log \text{Exp})] \\ &+ v_{it} - u_{it} \qquad (23)\end{aligned}$$

then, in the interest of parsimony, dropped the last two terms. Hollingsworth and Wildman [9], in the interest of comparability, adopted the same functional form. One might question the truncation of the production function. In principle, the translog form is used to approximate an arbitrary underlying function, and dropping those terms would significantly degrade the approximation. We refit the full translog stochastic frontier model. With all terms included, we found for COMP, first, that the scale elasticity, $\partial \log \text{COMP}/\partial \log \text{Exp} + \partial \log \text{COMP}/\partial \log \text{Educ}$, was roughly 0.35, suggesting large diseconomies of scale. Second, the function is not monotonic in the inputs for all values – marginal products were negative for many observations. Thus, the function would not be concave. Therefore, a strictly orthodox interpretation of the model as conforming to an optimization of output in the presence of a neoclassical production function seems optimistic. A much looser interpretation of the relationship between the health outcomes and the inputs seems more appropriate. To maintain continuity of this strand of analysis in the literature, we have maintained the simpler relationship

$$\begin{aligned}\text{LogHealth} &= \alpha + \beta_1 \log \text{Exp} + \beta_2 \log \text{Educ} \\ &\quad + \beta_{11} [(\log^2 \text{Educ})/2] + \cdots + v_{it} - u_{it} \\ &= \alpha + x'_{it}\boldsymbol{\beta} + \cdots + v_{it} = -u_{it} \qquad (24)\end{aligned}$$

Random and fixed effects estimates of inefficiency

Tables 5(a) and 5(b) present estimated production functions based on the simple panel data specifications, ETML's [4,5] (Cornwell *et al.*'s [3]) fixed effects model, the Pitt and Lee [24] random effects model, and the true fixed and random effects models.[m] None of these has any built in accommodation for cross country heterogeneity in the production function or the inefficiency. The estimated models differ somewhat, though the differences in the estimates of the structural parameters might be misleading – the estimated inefficiency terms (estimates of u_i) are quite similar for the two pairs of specifications, fixed/random effects and true fixed/true random effects. The estimated random effects models in Table 5(a) are consistent with Gravelle *et al.*'s [6,7] observation, that there is little within group variation. The variance decomposition is dominated by u_i; the estimates of $\lambda = \sigma_u/\sigma_v$ are 25.47 for DALE and 30.09 for COMP. Both of these are quite large by common standards. These values fall to 3.26 and 15.44 in the true random effects models for DALE and COMP, respectively. The relatively smaller values of 5.526 for DALE and 6.136 for COMP for the true fixed effects models appear consistent with this as well (though there is no counterpart in the ETML fixed effects model). The estimates suggest that these models in Table 5(b) are moving some of the variation out of u_i. This would be consistent with purging the invariant u_i of some time invariant heterogeneity. (We note, however,

Table 5.

		Constant	Expenditure	Education	Education^2	
(a) *Estimated frontier production models with time invariant inefficiency*[a]						
DALE	Fixed effects		0.00884	0.0629	0.0435	$s = 0.109$
			(0.00305)*	(0.0363)	(0.0285)	$R^2 = 0.998$
	Random effects	3.96	0.0152	0.0970	0.0130	$s_u = 0.2771$
		(0.0159)*	(0.00234)*	(0.0175)*	(0.0149)	$s_v = 0.0109$
						$\lambda = 25.47$
COMP	Fixed effects		0.00654	0.0495	0.0455	$s = 0.0066$
			(0.00185)*	(0.0221)*	(0.0171)*	$R^2 = 0.999$
	Random effects	4.24	0.0105	0.0617	0.0411	$s_u = 0.1989$
		(0.0101)*	(0.0011)*	(0.0101)*	(0.0925)*	$s_v = 0.00661$
						$\lambda = 30.09$
(b) *Estimated frontier production models with time varying inefficiency*[a]						
DALE	True fixed effects		0.0665	0.2896	−0.1147	$\lambda = 5.526$
			(0.00156)*	(0.0102)*	(0.0069)*	$\sigma = 0.2106$
						$\sigma_u = 0.2072$
						$\sigma_v = 0.0375$
	True random effects	3.632	0.04099	0.0172	0.1015	$\lambda = 3.26$, $\sigma = 0.0194$
		(0.0044)*	(0.0007)*	(0.0039)*	(0.0042)*	$\sigma_u = 0.01858$
						$\sigma_v = 0.0057$
						$\sigma_w = 0.1760$
COMP	True fixed effects		0.0684	0.1301	−0.0311	$\lambda = 6.136$, $\sigma = 0.1224$
			(0.00086)*	(0.0053)*	(0.0037)*	$\sigma_u = 0.1207$
						$\sigma_v = 0.0197$
	True random effects	3.551	0.0698	0.4762	−0.2559	$\lambda = 15.44$, $\sigma = 0.0781$
		(0.0121)*	(0.0025)*	(0.0185)*	(0.0185)*	$\sigma_u = 0.07093$
						$\sigma_v = 0.00459$
						$\sigma_w = 0.0773$

All variables in logarithms. Estimated standard errors in parentheses.
*Indicates significant at the 95% level.
[a] Based on 140 countries observed in all 5 years and Algeria observed in four. $\Sigma_{Ti} = 704$.

that the crucial difference might be in the means, rather than the variances. This is addressed below and in Table 6.)

Estimated inefficiencies are computed using the methods discussed earlier. In all cases, to simplify comparisons, we have used the direct estimate of inefficiency, $\hat{u}_i$ rather than $\exp(-\hat{u}_i)$. For the fixed effects estimator, this is simply $\max(a_i) - a_i$. The random effects estimator is computed using the Jondrow *et al*. [12] estimator in (7), as are the estimates for the true fixed and random effects models in Table 5(b). Tables 6(a) and 6(b) present analyses of these effects. As evidenced in the plots in Figure 1, the correspondence between the two sets of estimates, for both health outcomes is striking for both modeling approaches. The simple correlations between the pairs of estimates is almost 1 for the time invariant estimates, and 0.9515 and 0.8276 for DALE and COMP for the time varying estimates based on the 1997 observations.[n] The random effects estimators also reproduce the rankings of the fixed effects estimator which, in turn, gives largely the same results as ETML obtained with their normalized version. (France remains fourth in the DALE results and first in the COMP results, for example.) This degree of correspondence between these two estimators has been observed elsewhere (e.g., [34]). This finding suggests that the impact of the specific distributional assumption of the stochastic frontier model is not so severe as suggested earlier. The correspondence is similar between the true

Table 6.

	DALE		COMP	
	Fixed effects	Random effects	Fixed effects	Random effects
(a) *Analysis of estimated time invariant technical inefficiencies*				
Mean	0.2200	0.2089	0.1741	0.1647
Standard dev.	0.1862	0.1822	0.1181	0.1116
Correlation	0.9989		0.9988	
Second step regression results				
Constant	1.14 (0.156)*	1.075 (0.155)*	0.6965 (0.084)*	0.597 (0.0858)*
Gini	0.528 (0.139)*	0.546 (0.139)*	0.3084 (0.077)*	0.299 (0.0767)*
Voice	−0.0072 (0.0190)	−0.0089 (0.0189)	−0.0134 (0.010)	−0.0125 (0.0104)
GEFF	0.0352 (0.0230)	0.0365 (0.0229)	−0.0038 (0.0127)	−0.00451 (0.012)
LogIncome	−0.150 (0.0168)*	−0.144 (0.0167)*	−0.0779 (0.009)*	−0.0707 (0.009)*
Public finance	0.0009 (0.0062)	0.0008 (0.0006)	−0.0002 (0.0003)	0.0002 (0.0003)
OECD	0.0704 (0.0398)	0.0691 (0.0396)	0.0341 (0.0220)	0.0334 (0.0190)
R^2	0.5599	0.5447	0.6671	0.6292
(b) *Analysis of estimated 1997 observations on time varying technical inefficiencies*				
Mean	0.0797	0.0159	0.0462	0.0384
Standard dev.	0.0187	0.0171	0.0146	0.0190
Correlation	0.9515		0.8276	
Second step regression results				
Constant	0.1193 (0.222)*	0.0432 (0.0205)*	0.0665 (0.018)*	0.0975 (0.0207)*
Gini	0.0142 (0.0203)	0.0159 (0.0187)	0.0068 (0.016)	0.0319 (0.0189)
Voice	0.0006 (0.0027)	−0.0001 (0.0025)	0.0008 (0.0022)	0.0005 (0.0026)
GEFF	0.0017 (0.0033)	0.0009 (0.0031)	−0.00004 (0.002)	0.0016 (0.0031)
LogIncome	−0.0054 (0.0024)*	−0.0043 (0.0022)	−0.0024 (0.0019)	−0.0084 (0.0022)*
Public Finance	−0.00014 (0.001)	−0.0001 (0.0001)	−0.00001 (0.0001)	−0.0002 (0.0008)
OECD	−0.0005 (0.0058)	−0.00001 (0.005)	−0.002 (0.0046)	0.0002 (0.0054)
R^2	0.0729	0.657	0.0482	0.6371

Estimated standard errors in parentheses.
*Indicates statistical significance at 95% level.

random and true fixed effects values, though the correlation is somewhat reduced. Also to be expected, though somewhat disconcerting is the virtual lack of correlation between the CSS fixed and true fixed effects estimates, and between the Pitt and Lee [24] and the true random effects estimators. Not only are both the means and variances of the estimated inefficiencies much lower in the true effects models, but the full distribution over countries seems to have changed as well. But, again, this is consistent with the purging of some time invariant heterogeneity from the time invariant estimates in the CSS and Pitt and Lee [24] estimates. Figure 2 illustrates this for the comparison of the CSS fixed to the corresponding 1997 estimates from the true fixed effects model.

The lower panels of Tables 6(a) and 6(b) presents second step analyses of the estimated inefficiencies for each health measure for the two modeling approaches. They suggest that income and the distribution of income are both significant in explaining variation in efficiency when the inefficiency is assumed to be time invariant. Since u_i is in proportional terms, the absolute magnitudes of the coefficients give the proportional impacts. It appears that the most important determinant is the distribution of income, with larger values of the Gini coefficient (less equal income distribution) having a major negative impact on both health outcomes. (Increases in u_i imply lower efficiency.) The second largest determinant is per capita income, which works in the expected direction – higher income is associated with more efficient delivery of health care and achievement of higher life expectancy. (These results are not interpretable as direct impacts on

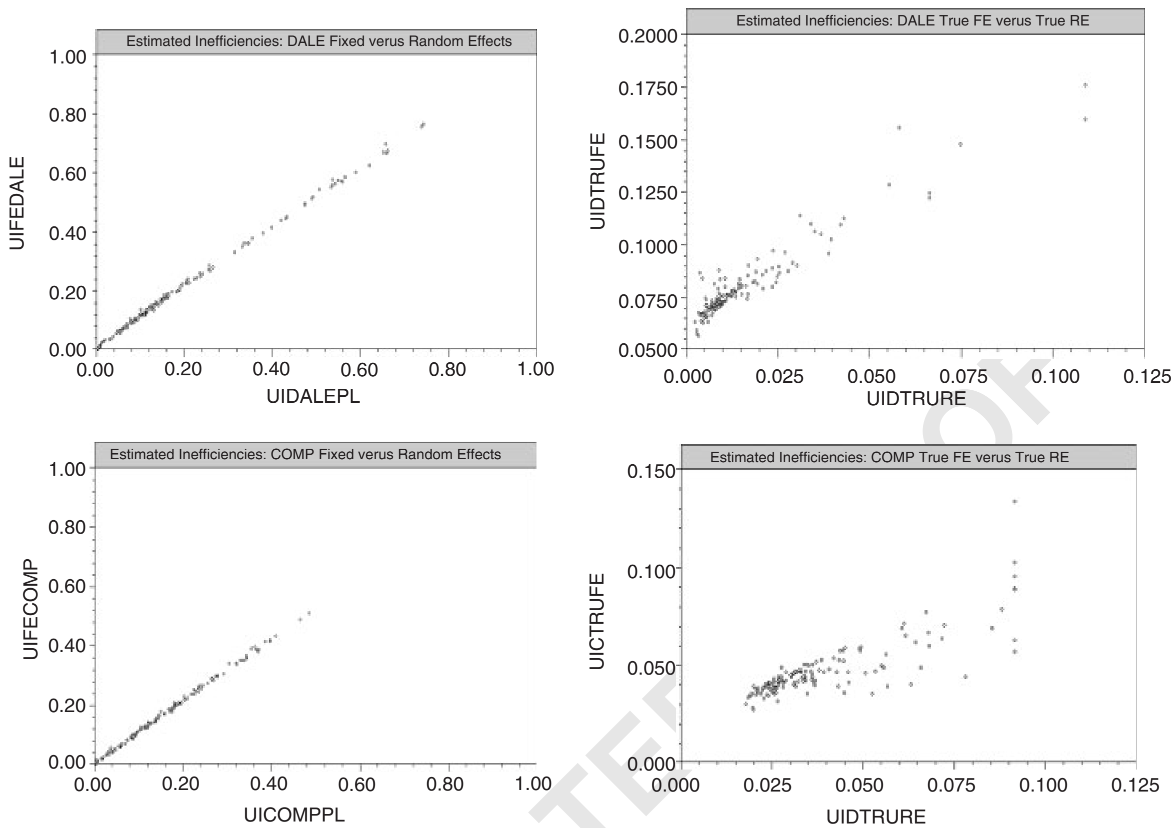

Figure 1. Estimated fixed and random effects inefficiencies (based on the 1997 observations for the time varying estimates)

the health outcomes.) Surprisingly, OECD status is associated with less efficient production. Note, though that per capita income is already in the equation, so whatever effect is at work in this persistent result is net of the impact of per capita income. Finally, as expected, this second step has very little explanatory power in the regressions involving the inefficiencies from the true random and fixed effects models.

The fixed effects estimators provides no method of accommodating time invariant indicators of heterogeneity in the model, but most of the others discussed do. The true random effects (random constant term) model, however, becomes highly unstable in these data when any other covariates are added to the model – this occurs because there is insufficient within group variation to estimate this type of model. Thus, from this point forward, we will focus on the random effects models. The true fixed and random effects approaches still seem promising, but analysis of them is held for further research.

Incorporating measured heterogeneity in the production function and in the estimates of inefficiency

We now consider expanding the functional form to accommodate the time effects and the other measured covariates. The augmented form of the production function will be at least

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\gamma} + \mathbf{t}'\boldsymbol{\theta} + v_{it} - u_i$$

where $\mathbf{z}_i$ = (TROPICS, POPDEN) – these two variables seem more appropriate as impacts on the level of production than as drivers of inefficiency. The placement of the remaining heterogeneity indicators, $\mathbf{h}_i$ = (GINI, Income,

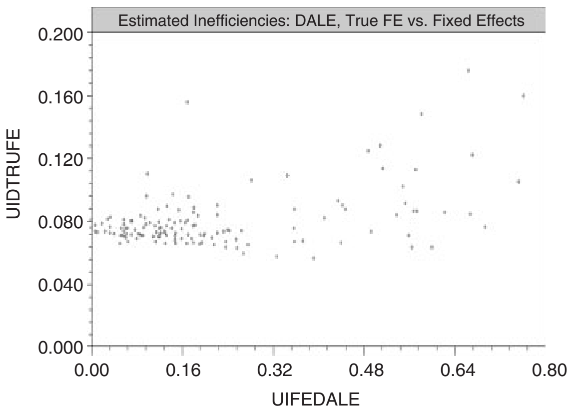

Figure 2. Estimated inefficiencies for DALE, True vs ETML fixed effects

OECD, Voice, GEFF, PUBFIN) remains to be established. There is no clearly defined theory that dictates how these should enter the model, so we approached the issue as follows: First, a practical result, when the GINI variable appeared in the production model, with or without the other elements of $\mathbf{h}_i$, the estimator produced wildly erratic estimates of the distribution of u_i – the estimate of σ_u in Table 5 of 0.2771 for DALE jumps to over 1.0, and the coefficient λ diverges to over 1000. The behavior of COMP is similar. When the truncation model is specified as the alternative, with

$$\mu_i = \delta_0 + \delta_1\ \mathrm{GINI}_i$$

the model behavior is quite reasonable, and shifts marginally from the results in Table 5. We concluded on this (admittedly nonstatistical) basis that GINI_i belonged more appropriately in the mean of the inefficiency distribution. The log of per capita income seems likewise to be more naturally associated with the efficiency of production than production, itself. Per capita expenditure already in the production function, will capture some of the effect of higher incomes in any event. Thus, our departure point is

$$\mu_i = \delta_0 + \delta_1\ \mathrm{GINI}_i + \delta_2\ \mathrm{LogIncome}$$

That leaves the set of indicators,

$$\mathbf{h}_i = \mathrm{GEff}_i, \mathrm{Voice}_i, \mathrm{PubFin}_i, \mathrm{OECD}_i$$

Though these would fit naturally in μ_i, they could arguably be drivers of production instead (or as well).

We have tested this specification issue as follows: We have two competing models. Model (P) places the indicators in the production function:

$$\text{(P)}\quad y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + \mathbf{t}'\boldsymbol{\theta} + h_i'\boldsymbol{\phi} + v_{it} - u_i$$

$$\mu_i = \delta_0 + \delta_1\ \mathrm{GINI}_i + \delta_2\ \mathrm{LogIncome}_i$$

while model (E) places them in the mean of the heterogeneity distribution;

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + \mathbf{t}'\boldsymbol{\theta} + v_{it} - u_i$$

$$\text{(E)}\quad \mu_i = \delta_0 + \delta_1\ \mathrm{GINI}_i + \delta_2\ \mathrm{LogIncome}_i + \mathbf{h}_i'\boldsymbol{\delta}_3.$$

These models are not nested, so no simple test (such as an F) test can be used to distinguish them statistically. We have used Vuong's [35] test for nonnested models, instead. The two competing models were estimated by maximum likelihood. Let L_i(P) be the contribution of country i to the log likelihood function assuming specification P and let L_i(E) be the counterpart for the other model. Then, let $q_i = L_i(\mathrm{P}) - L_i(\mathrm{E})$. The statistic is $\sqrt{n}\bar{q}/s_q$ where $\bar{q}$ and s_q are the sample mean and standard deviation. Vuong showed that this statistic has a limiting standard normal distribution. The test is directional. Values in excess of $+1.96$ (assuming a 95% significance level) favor model (P) while values less than -1.96 favor model (E). The intermediate values are (unfortunately) inconclusive. The estimated values of the Vuong statistic were 3.29 for DALE and 1.78 for COMP. These favor placing the indicators of heterogeneity in the production model – thus model (P) is our preferred specification.

Table 7 presents estimates of the random effects truncated normal stochastic frontier in which the time invariant covariates GINI and LogIncome appear in the underlying mean of U_i. The model also allows for technical change (the time shifts in the production function) and for some country effects in the production function (tropical location and population density and the other components discussed above). This is the heterogeneous form, with $E[U_i] = \mu_i = \delta_0 + \mathbf{h}_i'\boldsymbol{\delta}$.[o] The pattern in these results is similar to the preceding outcome. Once again, per capita income and the distribution of income appear to be significant determinants of the mean level of inefficiency, again in the expected direction.

The incorporation of cross country heterogeneity in the model has also produced the expected result with respect to the variation in inefficiency.

Table 7. Estimated heterogeneous truncated normal stochastic frontiers

		DALE		COMP	
		Truncation	Batt./coelli	Truncation	Batt./coelli
Production function	Constant	4.04	3.91	4.26	4.22
		(0.0319)*	(0.0236)*	(0.0209)*	(0.0197)*
	Exp	0.00826	0.00791	0.00609	0.00600
		(0.00306)*	(0.00215)*	(0.00146)*	(0.00145)*
	Educ	0.0484	0.241	0.0420	0.0916
		(0.0231)*	(0.0143)*	(0.0113)*	(0.0105)*
	Educ^2	−0.0146	−0.109	0.00498	−0.0224
		(0.0140)*	(0.0116)*	(0.00873)	(0.00863)*
	Year 1993	0.0300	0.00350	0.0178	0.0113
		(0.0134)*	(0.0138)	(0.0102)	(0.0113)*
	Year 1994	0.0320	0.00761	0.0191	0.0134
		(0.0147)*	(0.0156)	(0.0101)	(0.0129)
	Year 1995	0.0339	0.0120	0.0205	0.0156
		(0.0147)*	(0.0142)	(0.0104)*	(0.0129)
	Year 1996	0.0360	0.0166	0.0200	0.0179
		(0.0183)*	(0.0180)	(0.0120)	(0.0144)
	Tropics	−0.00979	−0.0237	−0.00591	0.00424
		(0.0136)	(0.0125)*	(0.0155)	(0.0142)*
	Pop. density	0.00557	−0.00309	0.00585	0.00503
		(0.00275)*	(0.00235)	(0.00192)*	(0.00197)
	Voice	0.0202	0.0184	0.0164	0.0189
		(0.00858)*	(0.00683)*	(0.0092)	(0.00893)*
	GEFF	0.00478	0.0143	0.0137	0.00926
		(0.00733)	(0.00775)	(0.0078)	(0.00742)
	Public finance	−0.0000896	0.000270	0.000117	0.000226
		(0.000302)	(0.000246)	(0.000229)	(0.000257)
	OECD	0.0383	0.0216	0.0237	0.0204
		(0.0117)*	(0.0131)	(0.00941)*	(0.0122)
Mean inefficiency	Constant	12.80	1.92	12.24	0.949
		(1.64)*	(0.435)*	(1.50)*	(0.116)*
	Gini	8.41	1.51	7.36	0.610
		(1.78)*	(0.456)*	(1.41)*	(0.117)*
	GDP Per capita	−1.98	−0.309	−1.68	−0.130
		(0.240)*	(0.0731)*	(0.176)*	(0.0150)*
Distributions of u and v	σ_u	0.162*	0.197*	0.0806*	0.0786*
	σ_v	0.0107*	0.00952*	0.00645*	0.00636
	λ	16.79*	20.72	12.51	12.36
	η		−0.0190*		−0.00699*
Log likelihood		1806.976	1861.687	2150.807	2158.697

Estimated standard errors in parentheses.
*Indicates statistical significance at 95% level.
[a] Sample is all 141 countries. $\Sigma_i T_i = 704$.

In Table 5, in the simple random effects model, the standard deviation of the underlying distribution of U_i, σ_u, is estimated as 0.2771 and 0.1989 for DALE and COMP, respectively. In the expanded model in Table 7, the counterparts are 0.1628 and 0.0806. A large proportion of the variation in 'inefficiency' appears to be explainable as heterogeneity in the mean, instead. The estimate of the residual variation, σ_v, is almost unchanged for both DALE and COMP.

The second set of estimates in each grouping in Table 7 are for Battese and Coelli's [26,27] extension of the random effects model. This form of the model incorporates some time variation in the inefficiency; the generic specification is

$$U_{it} = \exp[-\eta(t - T)]|U_i| \quad (25)$$

where t is the current year (1993–1997) and T is the terminal year (1997). Note that the Pitt and Lee [24] model is nested within this one, so we can test for the presence of this form of evolution of inefficiency. The log likelihoods are given in Table 7 for this purpose. For both variables, the hypothesis that η equals zero (that is, the hypothesis of the Pitt and Lee [24] model) is rejected. In practical terms, however, the scale factor in (25) brings only very minor change in the year to year estimates of u_i. The underlying random component of the inefficiency remains time invariant. The effect of this on the overall nature of the inefficiency is a bit ambiguous.

We have included in Table 8 descriptive statistics for the inefficiency estimates and comparisons to the fixed effects estimates produced by ETML. Figure 3 shows the correspondence graphically for DALE using the Pitt and Lee [24] random effects model. Virtually the same pattern arises for the Battese and Coelli [26,27] model, and for COMP using either model as well.[p] In fact, the statistical results in Table 7 fail to reveal some rather substantial adjustments. For many

Table 8. Estimates of technical inefficiency from heterogeneous random effects models

		DALE				COMP			
		Trucation		B&C		Trucation		B&C	
Estimated inefficiency (comparison to fixed effects estimates)	Corr. with FE	0.978		0.984		0.962		0.971	
	Rank corr.	0.932		0.945		0.962		0.968	
	Mean	RE	FE	RE	FE	RE	FE	RE	FE
	OECD	0.0607	0.0792	0.0639	0.0792	0.0554	0.0558	0.0381	0.0297
	NonOECD	0.231	0.258	0.221	0.258	0.181	0.206	0.173	0.206
	All	0.196	0.220	0.187	0.220	0.153	0.174	0.144	0.174
	Std. deviation	RE	FE	RE	FE	RE	FE	RE	FE
	OECD	0.0345	0.0474	0.0317	0.0473	0.0293	0.0400	0.119	0.112
	NonOECD	0.214	0.192	0.202	0.192	0.119	0.112	0.0297	0.0400
	All	0.204	0.186	0.191	0.186	0.119	0.118	0.120	0.118

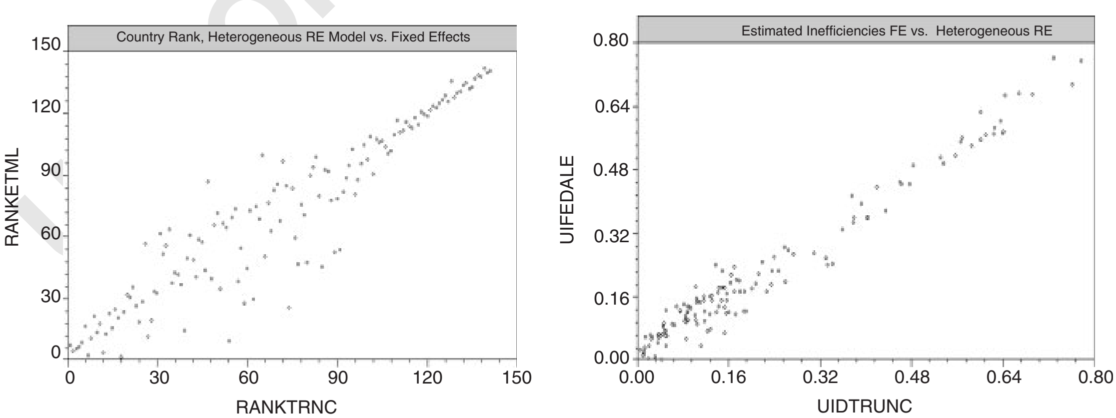

Figure 3. Country ranks and inefficiencies, DALE, random effects model

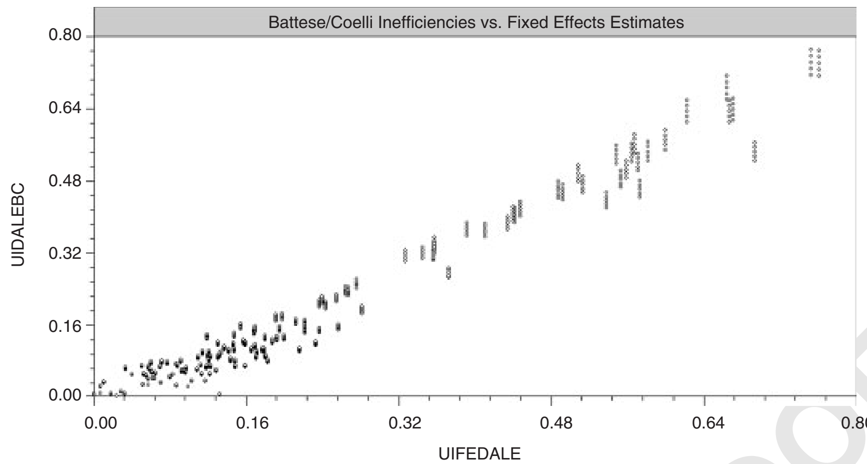

Figure 4. Battese/Coelli time varying inefficiency estimates

countries, this expansion of the model appears to be fine tuning, but for a large number of others, quite substantial differences emerge.

The results in Table 7 suggest that the efficiency of production is significantly affected by both the distribution of income and the level of per capita income, this persistent result has been hinted at in previous analyses. We do find considerable evidence of the result in our estimates. The highly significant estimate of η in the estimates of (25) (based on the likelihood ratio tests) suggest that the inefficiency varies somewhat over time as well. However, the plot in Figure 4 of the estimated values country by country reveals that this model extension actually adds relatively little actual variation in the estimates. The fixed effects estimate is the same in every year, so the vertical grouping reveals the time variation in the Battese and Coelli [26,27] estimates. Overall, 99.86% of the variation in the B&C estimates is between countries. Nearly 100% of the within group variation is accounted for by the 25 least efficient countries.

Figures 5(a) and 5(b) display the estimates of u_i from this model, once again in comparison to the initial estimates in ETML. We have segregated the estimates by OECD and non-OECD countries in the figures and in the descriptive statistics in Table 8. The dotted ovals in the figures contain all the OECD observations. This partition of the sample reveals a major aspect to all of the results. In the figures, the OECD countries are packed

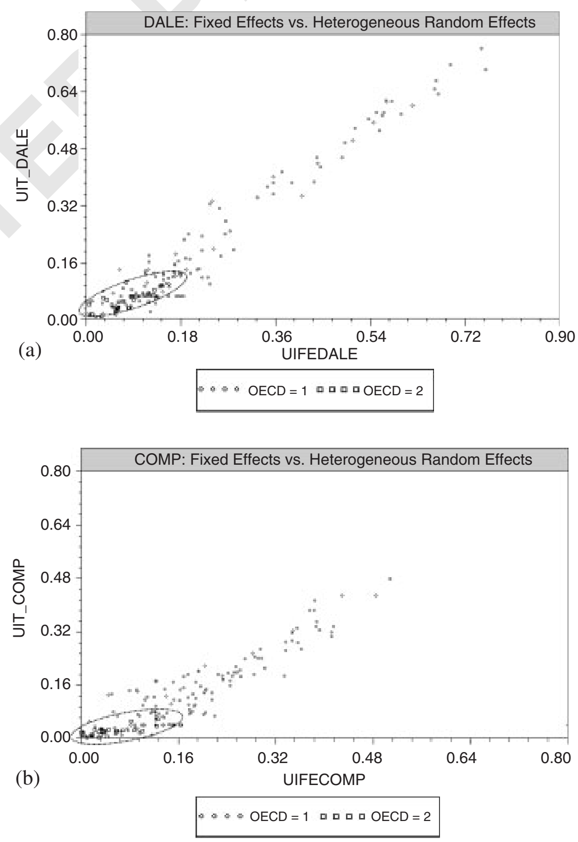

Figure 5. (a) DALE, fixed effects vs heterogeneous random effects, 1997 values, (b) COMP, fixed effects vs heterogeneous random effects, 1997 values

tightly in the lower left corner of the plots. From the descriptive statistics, as borne out by the figures, we now see that nearly all of the variation in the estimated inefficiencies in both ETML's fixed effects estimates, and these, our final proposed alternatives, is attributable to the non-OECD countries in the sample.

It can be seen in the figures that this model brings some rather large changes in the results. To explore this further, we have plotted the country ranks for the DALE estimates in Figure 6. (The figure for COMP is similar.) The results have changed considerably. Table 9 shows the effect of the respecified model on the ranks of the top 25 countries in the original fixed effects form. The table suggests that the rankings change considerably with this change in the model. Recalling that in the WHR, the rankings, not the absolute levels of inefficiency, were the primary object of estimation, we conclude that in the terms of the study, the model specification is clearly an important issue.

Conclusions

The WHO data analyzed here include most of the world's countries and embody nearly its entire population. The original studies by Evans *et al.*

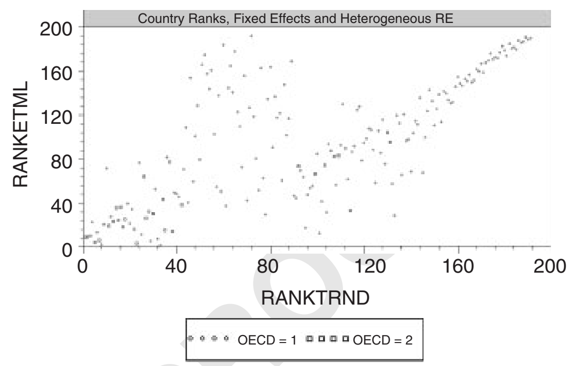

Figure 6. Country ranks: fixed effects and heterogeneous random effects, DALE

Table 9. Country ranks for the top 25 Countries in ETML sample

	DALE		COMP	
Rank	Country	New rank	Country	New rank
1	Malta	33	France	16
2	Oman	8	Italy	8
3	Italy	23	San Marino	19
4	France	5	Andorra	14
5	San Marino	29	Malta	10
6	Spain	7	Singapore	8
7	Andorra	32	Spain	2
8	Jamaica	2	Oman	90
9	Japan	1	Austria	9
10	Greece	3	Japan	4
11	Monaco	25	Norway	12
12	Saudi Arabia	101	Portugal	31
13	Singapore	6	Monaco	39
14	Portugal	38	Greece	1
15	Austria	35	Iceland	20
16	Norway	22	Netherlands	11
17	United Arab Emir.	95	Luxembourg	36
18	Netherlands	17	Ireland	33
19	Sweden	11	United Kingdom	22
20	Costa Rica	9	Colombia	29
21	Cyprus	20	Switzerland	24
22	Chile	4	Belgium	17
23	United Kingdom	13	Cyprus	21
24	Iceland	15	Sweden	3
25	Switzerland	18	Saudi Arabia	128

were an innovative, large scale effort to compare these countries on the effectiveness of delivery of health care using two measures, disability adjusted life expectancy (DALE) and a recently developed composite measure of health care delivery (COMP). The authors used a panel data technique based on the fixed effects regression model to obtain both quantitative measures of effectiveness and simple rankings for each country. Subsequent researchers have analyzed these same data with a wide range of agreement or disagreement with the methodologies and results. This study has continued that analysis, with several extensions. First, the model used here is considerably more general than those used previously. Second, we have incorporated measured indicators of cross country heterogeneity in the estimates. Third, we have produced an alternative set of individual country specific efficiency measures and country ranks. Our results differ substantially from those obtained by the authors of the original studies.

This reanalysis of the WHO data was motivated by several considerations.

- Both the fixed effects approach used earlier, and several others that have appeared in received studies make no distinction between technical inefficiency and cross country heterogeneity. Thus, some of what was reported as measured inefficiency should instead be treated as heterogeneity. Certainly, decomposing a two part unmeasured effect is a difficult exercise. The several models proposed here do so to varying degrees. We find, in our preferred specification, that making this distinction brings a substantial change in the estimated results
- The fixed effects model used in Evans *et al*. [4,5] does not allow the analyst to make use of measured effects that capture some of the cross country heterogeneity in the data. The models proposed here can accommodate these variables. We find including these measured indicators, such as per capita income and a measure of income distribution, produce noticeable changes in the estimated results.

In formulating the production function, we confirm what earlier researchers have found with respect to the income elasticity of health expenditure. Our estimates range from 1.08 to 1.23, with much larger values for the wealthier OECD countries than for the remaining countries in the sample. A number of earlier studies have examined the relationship between income (measured by per capita GDP) and health outcomes, and found a weak to nonexistent support. We find a persistent, significant impact of income on inefficiency. Likewise, the distribution of income has been suggested as an important influence, but prior results were somewhat inconclusive. As in the case of income, itself, we find significant explanatory power in the distribution of income. It should be noted that in both these cases, and throughout this study, the distinction between OECD and non-OECD countries, which explains much of the variation in these measures, also explains much of the variation in health outcomes and in the efficiency of delivery of them.

Using the random effects model as the platform, we found that comparing a basic model with no heterogeneity (Tables 5 and 6) to one with heterogeneity in both production and the mean of the inefficiency (Tables 7 and 8), the estimated underlying standard deviation in the distribution of u_i, σ_u, falls from 0.2771 to 0.1620 for DALE and from 0.1989 to 0.0806 for COMP. The sample means of the estimated inefficiencies also fall slightly, from 0.2088 to 0.1960 for DALE and from 0.1647 to 0.1530 for COMP. These are consistent with the original conjecture, that unaccounted for heterogeneity was indeed showing up as inefficiency in the original model. This does not imply, however, that the rankings of the countries would be changed. But, as seen in Figure 5, the rankings do change, considerably.

It should be emphasized, the measures and rankings in the tables pertain to the efficiency of delivery of health outcomes, not to absolute levels of those outcomes. Our results do not comment on the levels of health attainment in countries contained in this data set.

Acknowledgements

Elements of this paper have been presented at the conference on 'Current Developments in Productivity and Efficiency Measurement', University of Georgia, October 25–26, 2002 and the Second Hellenic Workshop on Productivity and Efficiency Measurement in Patras Greece, May 30, 2003. It has also benefited from comments at the North American Productivity Workshop at Union College, June, 2002, the Asian Conference on Efficiency and Productivity in Taipei in July, 2002 and from discussions at The University of

Leicester, York University and Binghamton University and ongoing conversations with Mike Tsionas, Subal Kumbhakar and Knox Lovell.

Notes

a. The participants were, in addition to Evans *et al*., William Greene of NYU, Subal Kumbhakar, University of Binghamton, Knox Lovell, University of Georgia, Kaliappa Kalirajan, ANU, Marijn Berhoeven, IMF, Paul Wilson, University of Texas, Christopher Tong, Hong Kong Baptist University and Philip Grossman, St. Cloud State University.

b. We have focused on parametric (stochastic frontier) and semiparametric (fixed effects linear regression) models. Nonparametric methods such as data envelopment analysis (DEA) are not considered here (for commentary, see, e.g. [9]).

c. Other distributional assumptions have been suggested, such as the normal-exponential [10] and the normal-gamma [15,16]. These extensions occasionally bring noticeable changes in the results. But they are tangential for present purposes.

d. The only received application of a type of true fixed effects model in the frontiers literature is Polachek and Yoon's [17] study of labor supply.

e. The major *practical* obstacle to use of the fixed effects approach in nonlinear models such as this one is the difficulty of computing the possibly hundreds or thousands of dummy variable coefficients. Unlike the linear regression model, the stochastic frontier model cannot be transformed to eliminate the dummy variables from the estimator. It is widely held that this renders the model unusable if the number of individuals (N) is very large (see, e.g., [18]), but see Greene [19] for analysis of the solution to this computational problem and Greene [22] for an application involving direct computation of some extremely large models. The *methodological* shortcoming of the fixed effects approach is the so called incidental parameters problem. See [20] and Greene (2004) for discussion. This issue as it relates to the stochastic frontier model remains completely unresolved; the only received evidence on it is the Monte Carlo study in Greene (2004).

f. In principle, one might estimate u_i in a random effects linear regression, either based on a full normality assumption, or as the MMSE projection, using $\hat{u}_i = E[u_i \mid \bar{\varepsilon}_i] = \rho\bar{\varepsilon}_i$ where $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$. However, this is still unsigned, and a second transformation to $\hat{u}_i^{\text{eff}} = \max[\hat{u}_i] - \hat{u}_i$ would still be needed. The usefulness of this estimator seems dubious. We are not aware of any applications of investigations.

g. This extension of the model introduces a subtle complication. The 'heterogeneity' introduced into u_{it} by this formulation influences both the mean and variance of inefficiency, because although η_{it} represents a scaling of the inefficiency, this scaling (the standard deviation) enters the mean as well, as can be seen in (15)–(16). More generally, stochastic frontier models that seek to incorporate *heteroscedasticity* in the distribution of u_{it} actually, because of this result, also build heterogeneity in the mean in to the model at the same time. Unlike the linear regression model, in the stochastic frontier model, the conditional mean of the part of the disturbance that is of interest is a function of both the underlying mean and the underlying variance.

h. The assistance of researchers at WHO, especially David Evans and Ajay Tandon is gratefully acknowledged. The data set used here is an expanded version of the data used in ETML [4,5]. The earlier papers did not use the invariant measures, per capita GDP, income distribution, etc.

i. One country, Algeria, was only observed four times.

j. Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, The Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States.

k. Certainly other country specific covariates could explain the heterogeneity in this data set. Those mentioned were contained in the larger data set. Observers might differ on the *best* set of variables. We leave it for ongoing research to fill out the model with an *optimal* treatment of the unobserved cross country heterogeneity.

l. The year 1997 dummy variable is the omitted category.
m. To avoid fitting panel models with 'groups' of one observation, the models to follow are based on the 140 countries with five years of data plus Algeria, with four.
n. Using the country means of the inefficiencies produces largely the same results.
o. The strikingly large constant and coefficient on GINI in the first and third columns do not indicate extremely large means, as they are offset by the coefficient on the log of GDP. The means of GINI and logGDP are 0.34 and 8.4, respectively, so it can be seen that these effects are largely offsetting in $E[U_i]$.
p. For the Battese and Coelli formulation, we did the comparison based on the 1997 estimate of u_i. There is a small amount time variation in the estimates from this model, but the patterns are essentially unchanged.

References

1. World Health Organization. *Health Systems: Improving Performance*. The World Health Report, 2000, Geneva. 2000.
2. Schmidt P, Sickles R. Production frontiers with panel data. *J Business Econ Statist* 1984; **2**: 367–374.
3. Cornwell C, Schmidt P, Sickles R. Production frontiers with cross sectional and time series variation in efficiency levels. *J Econ* 1990; **46**: 185–200.
4. Evans D, Tandon A, Murray C, Lauer J. *The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries*. World Health Organization, GPE Discussion Paper, No. 29, EIP/GPE/EQC, 2000.
5. Evans D, Tandon A, Murray C, Lauer J. *Measuring Overall Health System Performance for 191 Countries*. World Health Organization, GPE Discussion Paper, No. 30, EIP/GPE/EQC, 2000.
6. Gravelle H, Jacobs R, Jones A, Street A. *Comparing the Efficiency of National Health Systems: Econometric Analysis should be handled with care*. University of York, Health Economics, UK. Manuscript, 2002.
7. Gravelle H, Jacobs R, Jones A, Street A. *Comparing the Efficiency of National Health Systems: A Sensitivity Approach*. University of York, Health Economics, Manuscript, UK, 2002.
8. Williams A. Science of marketing at WHO? A commentary on World Health 2000. *Health Econ* 2001; **10**: 93–100.
9. Hollingsworth J, Wildman B. The efficiency of health production: re-estimating the WHO panel data using parametric and nonparametric approaches to provide additional information. *Health Econ* 2002; **11**: 1–11.
10. Aigner D, Lovell K, Schmidt P. Formulation and estimation of stochastic frontier function models. *J Econ* 1977; **6**: 21–37.
11. Kumbhakar S, Lovell K. *Stochastic Frontier Analysis*. Cambridge University Press: Cambridge, 2000.
12. Jondrow J, Materov I, Lovell K, Schmidt P. On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econ* 1982; **19**: 233–238.
13. Stevenson R. Likelihood functions for generalized stochastic frontier functions. *J Econ* 1980; **13**: 57–66.
14. Wang H, Schmidt P. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J Productivity Anal* 2002; **18**: 129–144.
15. Greene W. A gamma distributed stochastic frontier model. *J Econ* 1990; **46**: 141–164.
16. Greene W. Simulated maximum likelihood estimation of the normal-gamma stochastic frontier model. *J Productivity Anal* 2003; **19**: 179–190.
17. Polachek S, Yoon B. Panel estimates of a two-tiered earnings function. *J Appl Econ* 1996; **11**: 169–178.
18. Maddala G. Limited dependent variable models using panel data. *J Human Res* 1987; **22**: 307–338.
19. Greene W. *Fixed and Random Effects in Nonlinear Models*. Stern School of Business, Department of Economics, Working Paper 01-01, 2001.
20. Lancaster, T. The incidental parameters problem since 1948. *J Econ* 2000; **95**: 391–414.
21. Hsiao C. Logit and probit models. In *The Econometrics of Panel Data: Handbook of Theory and Applications*, Matyas L, Sevestre P (eds). 2nd Revised edn. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996.
22. Greene W. *Fixed Effects and the Incidental Parameters Problem in the Tobit Model*. Manuscript, Department of Economics, Stern School of Business, New York University, 2002b.
23. Greene, W. *Econometric Analysis*. Prentice-Hall: Englewood Cliffs, 2003c.
24. Pitt M, Lee L. The measurement and sources of technical inefficiency in Indonesian weaving industry. *J Dev Econ* 1981; **9**: 43–64.
25. Econometric Software, Inc. *LIMDEP Version 8.0*. ESI: New York, 2002.
26. Battese G, Coelli T. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *J Productivity Anal* 1992; **3**: 153–169.
27. Battese G, Coelli T. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Emp Econ* 1995; **20**: 325–332.

28. Kumbhakar S, Orea L. *Efficiency Measurement using a Stochastic Frontier Latent Class Model*. Efficiency Series Papers, University of Oviedo, 2002.
29. Greene W. *Estimation of a Latent Class Stochastic Frontier Model*. Manuscript, Department of Economics, Stern School of Business, New York University, 2003b.
30. Greene W. *Fixed and Random Effects in Stochastic Frontier Models*. Stern School of Business, Department Of Economics, Working Paper 02-16, 2002a.
31. Greene, W. *Distinguishing between Heterogeneity and Inefficiency*: *Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems*. Working Paper 03-10, Department of Economics, Stern School of Business, New York University, 2003d, (http://www.stern.nyu.edu/~wgreene/heterogeneityandinefficiency.pdf).
32. Berger A, Messer J. Public financing of health expenditures, insurance, and health outcomes. *Appl Econ* 2002; **34**: 2105–2113.
33. Newhouse J. Medical care expenditure: a cross national survey. *J Human Res* 1977; **12**: 115–124.
34. Berger A, Mester L. Inside the black box: what explains differences in the efficiencies of financial institutions? *J Banking Finance* 1997; **21**: 895–947.
35. Vuong Q. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 1989; **57**: 307–334.

