Regional variation in health care system performance: A dual-level efficiency approach applied to NHS pathology in England

ABSTRACT

Measuring regional imbalance in the performance of the health care system is of high importance to policy makers. There is a debate in the literature about how best to do this; despite well-defined micro and macro measures, there is uncertainty about which approach to take at the regional level. In this paper, the efficiency of the healthcare system is our performance metric. We apply a dual-level stochastic frontier (DLSF) (Smith and Wheat, 2012) to isolate inefficiency at two vertically distinct organisational levels: an upper level representing the effect of central management and policy; and a lower level representing the performance of individual production units, given that they have a degree of autonomy. In addition, we control for cross-unit heterogeneity, which remains an issue for healthcare performance analysis. We apply this approach to estimate regional performance in pathology services within the NHS. We use a panel on 57 pathology laboratories over a five year period. We find variation in performance at two organisational levels in pathology services: Strategic Health Authority (SHA) level (upper) and laboratory level (lower). We further compare our results to measures at a single-level and for which no control for heterogeneity is made to demonstrate the extent that these issues take effect on performance measures and thus underline the importance of accounting for these features.

1. INTRODUCTION

Performance analysis of health care systems can be conducted at one of three broad levels, macro (overall system), meso (regional) or micro (individual patients or providers) (Smith et al., 2012). Performance measures take many forms, from productivity and efficiency measures, to indicator measures such as in-hospital complications, or broader measures of public health such as life expectancy (Jacobs et al., 2006; Smith et al., 2012). In the NHS in England[1], meso-level, or regional, performance variation is an important aspect of health policy, and although there is little empirical analysis at this level (Bojke et al., 2013), two major policy drives underline its importance.

The first policy programme is the longstanding pursuit of health policy in England to reduce regional imbalance in resource allocation and, by extension, health inequalities (Marmot et al., 2010; Department of Health, 2007; RAWP, 1976)[2]. This policy is based on the Rawlsian principles of the

[1] Our focus is on the NHS in England in this study

[2] See http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4109403.pdf for a history of NHS resource allocation

NHS, stemming from the Beveridge Report on which they are based (Beveridge, 1942), aimed at promoting the welfare of society's least well off. Clearly, then, understanding these regional imbalances is necessary in pursuit of this goal.

The second policy pursuit is that, under the current financial strain, the NHS in England is looking to make efficiency savings, known as the Nicholson challenge, of £20bn by 2015 (Health Select Committee, 2010). Given that now 'easy savings' have been made (National Audit Office, 2012) and growing recognition that quality needs to improve alongside finances (Ham, 2013), there is growing scepticism that the Nicholson Challenge will be met (Appleby et al., 2013). One way in which to identify where savings can be made is to look for variation both in the performance across the system and at various organisational levels, such that areas performing away from best practice can be targeted for improvement.

Recent academic literature has begun to recognise that these organisational levels are not independent and therefore measuring performance of healthcare systems should take this into account (Sorensen et al., 2009; Castelli et al., 2013; Zhang et al., 2013). To this end, we examine the issue of regional variation by measuring inefficiency variation by region in pathology services in the NHS in England using a dual-level stochastic frontier (DLSF) (Smith & Wheat, 2012). We go a further step by decomposing the inefficiency at two vertically separate organisational levels. In doing so, we extend existing work in pathology performance evaluation (Buckell et al., 2013a; 2013b).

Our economic interpretation of the model, given the structure of pathology services in England, follows Smith and Wheat (2012): It is assumed that overall inefficiency comprises a combination of vertical inefficiencies, such that there is persistent inefficiency at the upper (SHA) level and residual inefficiency at the lower (laboratory) level.

First, then, is the upper component of inefficiency at the SHA level, derived from central management and control over pathology services (e.g. the configuration of services) in their region (Department of Health, 2006). This persistent inefficiency applies to all laboratories within that SHA. This upper component represents the performance of the i-th SHA relative to other SHAs in the NHS in England.

Second is the lower component of inefficiency which captures the variation of laboratories within the SHA. This represents the extent to which the particular laboratory fails to reach best practice within the same SHA.

Then, we are able to identify regional level inefficiency as a product of inefficiencies at the Strategic Health Authority (SHA) and at the laboratory (trust) levels.

Lastly, a Mundlak transformation is applied to ensure that inefficiency is measured net of purged unobserved heterogeneity (Mundlak, 1978; Farsi et al., 2005). We compare our results to a model which does not use such a device, to demonstrate the potential bias of unobservable heterogeneity.

The remainder of this paper is set out as follows. In section 2, the measurement of regional variation in health care system performance is discussed. Section 3 details the methods and data. Section 4 presents our results and discusses. Section 5 concludes.

2. REGIONAL PERFORMANCE IN HEALTHCARE SYSTEM MEASUREMENT

We consider three broad approaches in the academic literature to regional performance measurement in health care systems. The first is using aggregate measures per region; the second is studying providers at the micro level and averaging the results over each region to yield an estimate of performance measure; and the third is that which takes a multi-level approach looking to reconcile the two preceding approaches. Table 1 lists studies which have attempted to measure the regional performance of health systems[3]. There are a range of methods used including Data Envelopment Analysis (DEA), SFs, total factor productivity indices and multi-level models.

A central debate in the literature is the difficulty in measuring performance at the regional level. Bojke et al. (2013) argue that micro based studies are partial, or indeed may be subject to bias, since they typically focus on a single area of health care. On the other hand, results from aggregated, macro-level measures are sensitive to how they are constructed and therefore possibly unreliable (Black, 2012). Given the need to index outputs, it has been argued that results from aggregate measures are sensitive to their construction, both in the context of public sector performance measurement and in the wider economic literature (Theil, 1954; Blundell and Stoker, 2005; Smith and Street, 2005; Veillard et al., 2009).

Elsewhere, it has been argued that hierarchical models are well suited to systems of public service provision, in particular the NHS in England, and are thus useful analytical tools (Castelli et al., 2013). This may then represent a step forward in trying to assess regional variation in health care system performance (Sorensen et al., 2009). Multi-level models are becoming increasingly used in health care performance analysis for a number of applications to measure variation in health system performance (Olsen and Street, 2008; Castelli et al., 2013; Daidone and Street, 2013; Gutacker et al., 2013; Zhang et al., 2013).

[3] We exclude measures of cost variation from our review because there is little empirical analysis of regional variation in costs and spending (Bojke et al., 2013).

In this study, we address two issues with micro-based, multi-level models to yield a viable performance measure, before applying our measure to examine regional performance in pathology services.

First, when applied in the regional performance measure context, the performance estimate of the provider remains at a single level[4]. In the case where there are multiple organisational levels, the performance measure produced is likely to confound any persistent inefficiency at the upper level(s) with any varying inefficiency at the lower level(s). Thus, it is not clear to what extent the upper tier or lower tier is responsible for any sub-standard performance. Moreover, inefficiency is likely to be underestimated if its multi-level nature is ignored (Smith and Wheat, 2012). Therefore, our innovation is in the separation and then combination of these performance measures to estimate overall regional performance.

Second, as noted in Olsen and Street (2008) and in Sorensen et al. (2009), the multi-level measure used in these models is akin to a Schmidt and Sickles (1984) SF approach. The estimated (cost or production) function is, of course, only an approximation to the true function. As such, there are likely other differences between firms than inefficiency. Thus inefficiency is vulnerable to bias from unobservable heterogeneity if it not elsewhere captured in the model (Greene, 2004). The resolution to this is to separate out the inefficiency from the unobserved heterogeneity.

Therefore, we extend a DLSF using the Mundlak transformation to decompose inefficiency at different organisational levels and to address unobserved heterogeneity (Farsi et al., 2005). In addition, we statistically test the transformation and demonstrate the potential bias of its omission.

[4] However, recent studies (Zhang et al., 2013) have shown the potential for a greater number of levels in the data and so in the performance estimates.

Year	*Country*	*Author*	*Journal*	*Method*	*Organisation Level*	*Key Points*
2000	USA	Pai et al	J of Med Sys	DEA	Micro	Primary Care; sinusitis condition; average performance of units across regions
2000	USA	Scheffler et al.	Inquiry - Blue Cross and Blue Sheild Assoc.	Random Effects OLS	Micro	Mental health; before-after analysis (coefficient on control period); variation in performance present before and after intervention
2002	Switzerland	Crivelli et al.	J of Health Care Finance and Economics	SFA	Micro	Nursing homes; cost frontier; cross-section; cantonal dummies in cost function - not efficiency per se
2003	Botswana	Ramanatha n et al.	Development Southern Africa	DEA & SFA	Regional	Health districts; regional -level analysis on districts; very little variation in efficiency based on DEA outputs (20/22 districts fully efficient)
2006	Ukraine	Pilyavsky et al.	Health Economics	DEA	Micro	Hospitals; two stage - 1. DEA, 2. Tobit; western health system quicker to improve efficiency than eastern counterpart
2007	Mexico & Argentina	Jayasuriya & Wodon	Estudios Economicos	SFA	Regional	Production function approach; substantial variation across regions; degree of variation dependent on measure used
2007	China	Zhang et al.	Front. Econ. China	DEA	Regional	Provincial system as DMU
2008	Ukraine	Bernet et al.	JPA	DEA	Micro	Polyclinic analysis; scores averaged to compare between regions
2008	Ukraine	Pilyavsky et al.	JPA	DEA	Regional	Polyclinics and hospitals averaged over region; order-m DEA approach to estimate efficiency; Malmquist index for efficiency change over time
2008	Thailaind	Rajitkanok et al.	Health Care Man. Sci.	DEA	Micro	Hospital analysis; aggregated by region
2008	India	Purohit	RURDS	SFA	Regional	Analysis of regional output; production function approach; cross-section
2009	Portugal	Amado et al	Health Policy	DEA	Regional	Primary care; individual centres aggregated across regions
2009	USA	Lenard et al	Health Serv Man Res	DEA	Regional	Nursing homes; outputs of all homes in state combined, then state compared as DMU
2009	USA	Sikka et al	Health Care Mangt Rev	DEA	Regional	Hospitals; outputs combined into clusters; clusters taken as DMU
2009	Denmark	Sorensen et al	Health Policy	Multi-level	Micro- Regional	General practice; dual-level cost function; identifies differences in costs; inefficiency by Schmidt and Sickles FE approach; all composed inefficiency (OTE); cross section
2010	China	Gai et al	Biosci Trends	DEA	Micro	Hospital analysis; averaged over region
2010	Canada	Law et al	Can J of Reg Sci	DEA	Micro	Patient-level analysis; cancer treatment; averaged over regions
2010	India	Purohit	Soc Work Public Health	SFA	Regional	Analysis of regional output; production function approach; cross-section
2010	Brazil	Varela et al	Health Care Man Sci	DEA	Regional	Analysis of regional output by municipality; cross-section

2010	Spain	Zafra-Gomez et al	Env Plan C: Gov and Pol	DEA	Regional	Local authority as DMU; two years of data used
2011	Pakistan	Abbas et al	J Econ and Beh Stud	DEA	Regional	Basic Health Units as DMU
2011	Greece	Halkos et al	Health Policy	DEA	Regional	Prefecture (region) as DMU
2011	Greece	Kontodimopoulos et al	J Med Sys	DEA & SFA	Micro	Dialysis facilities; 2nd stage regression included region
2011	Spain	Murillo-Zamorano et al	Eur J Health Econ	SFA	Micro	Primary care; average across regions; cross section; production function approach
2011	China	Ng	China Econ Rev	DEA	Micro	Hospital analysis; averaged over region; panel 5 years
2012	England	D'Amico et al	App Econ Pers Pol	SFA	Regional	Social services; local authority level; panel 6 years; cost function approach
2012	China	Hu et al	China Econ Rev	DEA	Micro	Hospital analysis; averaged over region; panel 6 years
2012	Italy	Pelone et al	Health Policy	DEA	Regional	Primary care; regions as DMUs
2013	England	Bojke et al	Health Economics	TFP Index	Regional	SHA regions analysed; productivity indices
2013	Germany	Felder et al	Eur J Health Econ	DEA	Regional	Region as DMU; order-m efficiency model
2013	England	Castelli et al	Soc Sci & Med	Multi-level	Micro-Regional	Variance of indicator measure as performance metric; estimated variance at separate organisational levels by health system structure and political system structure

Table 1: Measures of regional inefficiency and productivity

3. METHODS

3.1 Dual-level stochastic frontier model

Our principle model is the dual-level stochastic frontier as proposed by Smith and Wheat (2012), which we adapt for our application to pathology services. Laboratory costs are our metric of interest.

$$\ln C_{itl} = \alpha + f(X_{itl};\beta) + u_{itl} + v_{itl} \quad (1)$$

Where i = 1, …, N, N=10; t = 1, …, T(i), T=5; l = 1, …, L(i); L=57. C_{itl} is the cost of laboratory l in SHA i in time period t. α is a constant, X_{itl} is a vector of (logged) outputs, input prices and environmental variables, β is a vector of parameters to be estimated. v_{itl} is a random variable representing statistical noise and u_{itl} is a variable representing inefficiency.

So as to separate inefficiency to two levels, the u_{itl} is decomposed,

$$u_{itl} = \mu_{it} + \tau_{itl};\ \mu_{it} \perp v_{itl}, \mu_{itl} \sim iid, \tau_{itl} \sim iid \quad (2)$$

Here, u_{itl} is decomposed into two elements: μ_{it} which is a persistent element of inefficiency, attributable to SHA i, which applies to all laboratories within that SHA; and τ_{itl} which is the residual component of inefficiency which varies randomly across laboratories. These inefficiencies may be fixed or vary over time (as above).

Finally, given these two measures, it is then necessary to compute an overall inefficiency for the SHA – our regional performance measure - which is the sum of its persistent inefficiency and the (cost) weighted average of its constituent laboratories, as below.

$$\bar{u}_{it} = \mu_{it} + \frac{\sum_{\forall l} C_{itl} \cdot \tau_{itl}}{\sum_{\forall l} C_{itl}} \quad (3)$$

3.2 Mundlak transformation

One approach using standard panel data models, following Farsi et al. (2005), is to make the assumption that the unobserved heterogeneity is correlated with the regressors[5][6], but inefficiency is not. In this case, it is possible to partition a fixed effect into two parts.

[5] See Farsi et al. (2005), for discussion of this assumption

[6] We are aware that there are several other approaches to deal with unobservable heterogeneity in the literature. Examples include approaches for cross-sectional dependence in panels such as residual multifactor, spatial and common correlated effects approaches (Pesaran and Tosetti, 2011); and direct approaches such as 'True' random and fixed effects SFs (Greene, 2005), latent class SFs (Orea and

$$\alpha_{it} = X_{itl}'\rho + \delta_{itl}\ ;\ \delta_{itl} \sim iid\left(0, \sigma_{\delta}^{2}\right) \quad (4)$$

The first term, X_{itl}'ρ, captures the correlation between the fixed effect, α_i, and the regressors, representing the unobserved heterogeneity. The residual term, δ_{it}, is a random variable orthogonal to the regressors, capturing the persistent inefficiency of the SHA. Noting that both α_{it} and δ_{it} are laboratory invariant, this relationship is robust to averaging over individual laboratories.

$$\alpha_{it} = \bar{X}_{it}'\rho + \delta_{it}\ ;\ \delta_{it} \sim iid\left(0, \sigma_{\delta}^{2}\right) \quad (5)$$

Equation (5) has become known as the Mundlak transformation (Farsi et al., 2005).

3.3 Model

We substitute (5) into the cost frontier equation (1) to yield the Mundlak-transformed DLSF.

$$\ln C_{itl} = \bar{X}_{it}'\rho + \delta_{it} + f(X_{itl};\beta) + \tau_{itl} + v_{itl}\ ;$$

$$\delta_{it} \sim iid\left(0, \sigma_{\delta}^{2}\right), \tau_{itl} \sim iid|N(0, \sigma_{\tau}^{2})|, v_{itl} \sim iid\ N(0, \sigma_{v}^{2})\ ;\ \ \delta_{it} \perp \tau_{itl} \perp v_{itl} \quad (6)$$

Thus, model (1) has *k* additional regressors (k being the number of X), with the inefficiency components being random variables. The model is estimated as a RE GLS and then a second stage maximum likelihood frontier. Importantly (and the fundamental result of Mundlak's paper), the estimates of β using this approach are identical to the fixed effects approach (and so are consistent) and the GLS RE is a more efficient estimator.

3.4 Data and estimation

Annual pathology benchmarking data (Keele Benchmarking) is used to compile an unbalanced panel of 57 English NHS pathology laboratories during a 5 year period from 2006/7 to 2010/11[7]. The sample represents approximately one third of the 163 NHS pathology laboratories in England. From table 2, there is considerable variation in the range and standard deviation of the costs, tests and requests variables, giving us confidence that we have a broad sample of laboratories.

Our data is for biochemistry services only. Biochemistry is one of five disciplines of pathology (the other four being haematology, hystocytology, immunology and microbiology). Biochemistry is

---

Kumbhakar, 2004) and SFs based on the closed-skew normal distribution (Columbi et al., 2011). However, we note that all of these approaches are best suited to large N, large T panels. Here, not only is the panel fairly small in T, it is unbalanced. Indeed, in Buckell et al. (2013a) the true random effects model was shown to be difficult to estimate on this data. We therefore do not see these approaches as appropriate for this data.

[7] Number of observations over time of laboratories: t=2: 27, t=3: 7, t=4: 2, t=5: 21

chosen because it is highly mechanised thus diminishing the issue of heterogeneity for modelling. It is the largest area of pathology (around 70% total activity (Holland et al., 2011)) and all laboratories run biochemistry services.

Variables include total operating costs (net of capital charges), output (through number of tests and number of requests), input prices of labour (from the UK labour force survey) and exogenous variables including the tests-to-requests ratio and dummy variables for the foundation status of the host trust, for the pathology service providing teaching and for the laboratory type (metropolitan, meaning within an urban area). Service quality is assumed given that laboratories have been accredited.

Costs and wage data are in real terms (2007 prices) using the consumer prices index. Labour force survey data is chosen over other sources (NHS staff census data, for example). This is firstly to ensure the exogeneity of the data[8]. Secondly we aim to better reflect the regional variation in labour input prices than would be possible using alternative data. The ratio of tests to requests is calculated from the data.

Our estimation strategy is as follows. We first ignore unobservable heterogeneity and compare a DLSF to a single-level SHA SF (i.e. a model which supresses the lower level) and a single-level laboratory SF (i.e. a model which supresses the upper level). We show that the single-level methods have different efficiency estimates to the DLSF. We use statistical tests to justify our use of the DLSF (in addition to our a priori justification of this model). Having done this, we then proceed to apply the Mundlak transform detailed above. We then derive our regional performance measure using production-weighted efficiency estimates. LIMDEP software is used for estimation (Greene, 2012a).

Table 3: Descriptive Statistics

Variable	*Mean*	*S.D.*	*Min*	*Max*
Operating costs (adjusted)	3617320	2058358	963875	11741895
Number of tests	5037362	2990846	1380384	30199502
Number of requests	714125	465535	191078	4423531
Input prices (Labour) (adjusted)	24551	4160	15834	49955
Number of primary care tests	2059689	932794	380790	5480395

[8] Mutter et al. (2013) demonstrate using healthcare data that endogeneity can bias efficiency scores.

4. RESULTS

4.1 DLSF results

Table 4: Parameters estimates from DLSF and comparator models.

	Dual-level SF (1st stage)	*Single-level SHA SF*	*Single-level laboratory SF*
Variable			
Constant	(-)5.275 (1.74)***	(-)5.275 (1.74)***	2.315 (2.831)
OUTPUT	0.857 (0.043)***	0.857 (0.043)***	0.506 (0.043)***
INPUT PRICES	0.775 (0.157)***	0.775 (0.157)***	0.484 (0.289)*
TESTS:REQUESTS	0.520 (0.070)***	0.520 (0.070)***	0.284 (0.054)***
TIME	0.009 (0.013)	0.009 (0.013)	0.017 (0.010)*
METROPOLITAN	0.193 (0.050)***	0.193 (0.050)***	0.202 (0.111)*
FOUNDATION	(-)0.084 (0.044)*	(-)0.084 (0.044)*	(-) 0.055 (0.082)
TEACHING	0.013 (0.042)	0.013 (0.042)	0.079 (0.078)
Moulton-Randolph	3.11	3.11	
Lambda			3.24***
Lambda (2nd stage)	1.58***		

*, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively. Standard errors are in parentheses.

Table 4 shows the parameter estimates from the DLSF and the single-level SF. Note that the single level SHA SF is in fact the same model as the first stage DLSF. It can be seen that the results are similar across the models to a large extent. The output coefficients are all positive and significant, however they do suggest slightly different returns to scale - the single level laboratory parameter suggests quite strong economies of scale properties, whereas the DLSF and SHA SF suggest something closer to exhausted economies of scale. Elsewhere, the parameter estimates are much more in line and similar to findings in other studies of pathology services (Buckell et al., 2013a; 2013b). The purpose of this study is not to discuss in detail the elasticities with respect to cost. However, one clear point to note is that, on the basis of the estimated parameters, the models appear to be similar.

The Moulton-Randolph statistic indicates that the SHA stratification is statistically significant (the score is > 1.645). This supports our use of the SHA structure as compared to a pooled model. The lambda values for both models are highly statistically significant, indicating the presence of inefficiency in the data for both the single level laboratory SF and the 2nd stage of the DLSF.

Table 5 shows the mean values for efficiency estimates from models at the two organisational levels separately and the DLSF. Here, there appears to be an underestimation of inefficiency in both of the single level models compared to the DLSF. This is in keeping with Smith and Wheat (2012)[9].

In addition, there is no indication as to what extent the laboratories, rather than the SHA within which they lie, are driving their estimated efficiencies. The problem for the regulator is the inability to isolate the source of the inefficiency, thus making reducing said inefficiency problematic. This is not the case with the DLSF.

Given these results and our introductory chapters, we prefer the DLSF for theoretical, practical and statistical reasons.

Table 5: Mean efficiency estimates of DLSF and comparator models

	Persistent CE SHA	*Mean CE LAB*	*Overall CE Region*
DLSF	0.795	0.809	0.643
SL SHA	0.795	n/a	0.795
SL LAB	n/a	0.699	0.699

4.2 Mundlak transformation of DLSF and regional performance measures

[9] We note, however, in the case of the single level SF, this result is somewhat forced, given that the second stage is not estimated. (Although, of course, there is no guarantee that inefficiency will be detected at all at the lower level in the DLSF.) This is not the case for the laboratory level SF.

Table 6: Parameter estimates from DLSF with Mundlak transform.

	Dual-level SF Mundlak	*Fixed Effects*	*Dual-level SF Mundlak without TEACHING*
Variable			
Constant	-16.826 (7.627)**		1.860 (5.190)
OUTPUT	0.852 (0.042)***	0.852 (0.045)***	0.856 (0.041)***
INPUT PRICES	0.910 (0.163)***	0.910 (0.175)***	0.901 (0.163)***
TESTS:REQUESTS	0.522 (0.067)***	0.522 (0.072)***	0.526 (0.066)***
TIME	0.007 (0.013)	0.007 (0.013)	0.008 (0.013)
METROPOLITAN	0.187 (0.048)***	0.187 (0.052)***	0.196 (0.047)***
FOUNDATION	-0.074 (0.043)*	-0.074 (0.046)	-0.065 (0.041)
TEACHING	0.029 (0.040)	0.029 (0.043)	
REQBAR	0.637 (0.346)*		-0.310 (0.185)*
INPBAR	-0.132 (0.338)		-0.332 (0.444)
TESBAR	0.812 (0.737)		-1.075 (0.545)**
YEABAR	0.095 (0.107)		0.238 (0.134)*
AREBAR	0.107 (0.160)		0.116 (0.203)
FOUBAR	-0.507 (0.276)*		0.204 (0.171)
TEABAR	1.151 (0.395)***		
Wu test (Wald Chi-squared)	29.57***		13.82**
Wald test of non-significant group mean variables	8.61*		2.45

*, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively. Standard errors are in parentheses.

Table 6 shows the parameter estimates from the DLSF with the Mundlak transform. These are identical to a comparable fixed effects estimator, as shown, except that the standard errors are smaller for the RE. We have also included the Mundlak-transformed DLSF without the TEACHING variable.

Our reason for estimating a model without the TEACHING variable is its behaviour with the Mundlak transform applied. There is a puzzling situation whereby the variable itself is not statistically significant in the model which suggest no effect on costs based on within variation. However, its group mean is highly statistically significant.

In Farsi et al. (2005), significant group mean variables are deemed to represent association between the random effects and the variable. This is in keeping with the principle of the Mundlak transform, which is to filter out the unobservable heterogeneity from the inefficiency, rather than to be itself explaining costs. If this is the case, it would be intuitively reasonable to expect that teaching would be associated with a higher amount of unobservable heterogeneity. We note, however, that there are

measurement issues with this variable (see Buckell et al., 2013b) that may be causing some questionable statistical behaviour.

This notwithstanding, an important corollary is that the average SHA cost efficiency declines from around 0.972 (see table 7) to around 0.918 in the case of the DLSF Mundlak without teaching, implying that there is a significant effect of including or excluding this variable. Clearly, model selection has implications for the conclusions of the study. As can be seen, the results for the remaining variables are not significantly affected with its removal.

The Wald tests of 7 linear restrictions – the Wu test (Baltagi, 2008) - indicate that the variable means are jointly statistically significant additions to the models, suggesting that the Mundlak transformation is appropriate on statistical grounds[10]. We have also tested the group mean variables that are not statistically significant above and found that they are jointly significant additions to the model with TEACHING and not to the model without TEACHING. On the basis of this result, we opt for the model which includes the TEACHING variable.

Table 7: Model cost efficiency estimates. CE – cost efficiency, s.e. – standard error

SHA	*Dual-level SF (Mundlak)*			*Dual-level SF*		
	SHA CE	Laboratory CE (weighted)	Regional CE	SHA CE	Laboratory CE (weighted)	Regional CE
1	0.973	0.793	0.772	0.908	0.793	0.612
2	0.976	0.830	0.810	1.000	0.830	0.673
3	0.966	0.784	0.757	0.795	0.784	0.594
4	0.974	0.809	0.788	0.789	0.809	0.638
5	0.973	0.821	0.798	0.658	0.821	0.655
6	0.969	0.793	0.769	0.714	0.793	0.610
7	0.968	0.832	0.806	0.719	0.832	0.671
8	0.952	0.774	0.737	0.808	0.774	0.570
9	0.969	0.823	0.797	0.780	0.823	0.656
10	1.000	0.784	0.784	0.781	0.784	0.615
Avg.	0.972	0.804	0.781	0.795	0.804	0.629
s.e.	0.012	0.021	0.023	0.098	0.021	0.034

[10] We have also used the more familiar Hausman test. In this case, however, the test statistic could not be computed because the variance-covariance matrix is not positive definite. We thus revert to the Wu test (Greene, 2012b) and note that, in any case, reliance on the Hausman statistic alone is discouraged (Baltagi, 2008).

Table 7 shows the (relative, not absolute) cost efficiency estimates from the DLSF models estimated with and without the Mundlak transform applied. The upper level cost efficiency estimate is given in the first column, SHA CE; the weighted lower level estimates in the second column, laboratory CE – the average across all laboratories within each SHA; and the regional performance estimate, the product of the two, is the third column, regional CE. We do not comment on which SHA corresponds to which region to maintain anonymity.

For the DLSF with Mundlak, cost efficiency is high at the upper level, ranging from 0.952 to 1. Laboratory averages range from 0.774 to 0.832, meaning regional cost efficiencies range from 0.737 to 0.810, with an average of 0.781. Fig. 1 is a graph showing unweighted individual laboratory cost efficiencies. The finding that variation in performance is mostly at the lower tier of the organisation has been found in other studies (Castelli et al., 2013).

For the DLSF without the Mundlak transform, there appears to be much more inefficiency at the upper level, with cost efficiencies ranging from 0.658 to 1. The findings are identical for the lower level cost efficiency estimates, yielding a range of 0.570 to 0.673 for regional cost efficiency, with an average of 0.629.

Standard errors for estimates are identical for the laboratory level estimates, so the overall variation in the regional CE estimates is driven directly by the variation in the upper level estimates. It is therefore unsurprising that the DLSF with Mundlak has much lower standard errors.

Comparing the two models, this result is not surprising, given that the DLSF does not control for unobservable heterogeneity, meaning that this is recorded as inefficiency in this model. If our assumption[11] is upheld, the difference between the two models' efficiency estimates is the (quantifiable) unobserved heterogeneity. This point has, of course, been made elsewhere in the health literature (Greene 2004, 2005; Farsi et al., 2005).

Our average lower tier (laboratory) CE estimates in table 7 have been weighted by output. We apply a weight because, when considering a regional measure, it is important to take into consideration not just the efficiency estimates of each production unit, but also the amount of production at a given level of efficiency. If this is overlooked, it might be argued that the measure does not truly reflect the performance in that region. Secondly, we weight our estimates to counteract the effect of SHA size, which is to cause the mean of SHA efficiency estimates to approach the underlying overall sample

[11] Unobserved heterogeneity is correlated with the regressors

mean[12]. Given this effect, we emphasise the importance of individual laboratory efficiency estimates for regulatory purposes. Indeed, average measures may mask outliers, which are, of course, of specific interest to regulators. We therefore report the unweighted efficiency estimates by laboratory within each SHA (fig. 1). Lastly, when calculating our reported potential cost savings we weight estimates by cost, as per section 3.

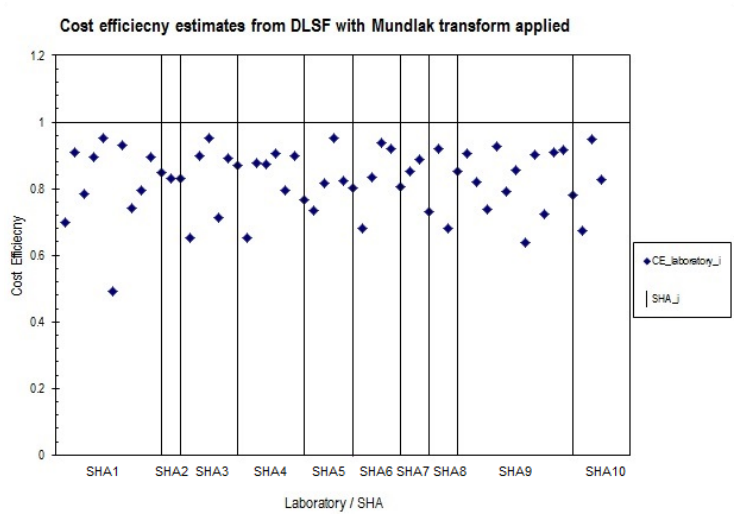

Fig. 1: Individual laboratory cost efficiency (unweighted) estimates from DLSF with Mundlak transform by SHA.

On average (table 7), it can be seen that SHAs 1, 3 and 8 should be targeted by the regulator for improvement. When the weighting is removed, as in figure 1, SHAs 1, 3 and 9 appear to house laboratories which have low efficiency estimates. However, given the results of both, there is no clear signal that one or several SHAs are performing significantly worse (or better) than the rest.

In terms of savings, we have found the potential for 22% efficiency gains. To calculate the potential monetary efficiency savings, we take the efficiency estimate of each laboratory in its final year, apply its cost weight and calculate the potential saving per laboratory. When this is aggregated across all of the laboratories, we find £48m of potential gains in the sample. If this is applied to all pathology services, this would suggest potential savings of around £600m. This is significantly more than found elsewhere (£250-500m in Department of Health, 2008; £450m in Buckell et al., 2013a). This follows from our findings: where the dual level structure is overlooked, inefficiency is underestimated and so here, a higher potential cost savings estimate should be yielded.

5. CONCLUSIONS

[12] We also found this property in our Pitt and Lee (1981) single level SFs and the Random effects model with the Schmidt and Sickles (1984) transformation applied.

We have estimated regional performance in pathology services in the NHS in England. We have found SHAs 1, 3 and 8 to be performing least well, and SHAs 2 and 7 best. Overall, we have found inefficiency in pathology services in England of around 22%. This would imply savings of approximately £600m if applied to all NHS pathology services.

We have introduced a DLSF to decompose inefficiency at two vertically separate organisational levels: at the upper SHA level, and at the lower laboratory level. We found that inefficiency resides mostly at the laboratory level (having weighted our efficiency estimates), and a small amount of inefficiency at the upper level. Importantly, from a regulatory perspective, we have specifically identified where the inefficiency resides within the system.

We have demonstrated, using two comparator models, that the estimated inefficiency is biased in two cases: one where the organisational structure of services is not fully accounted for; and another where unobservable heterogeneity is not separated out from the inefficiency. Given that the estimated parameters of these models are very similar, it appears that subtle differences can make a vast difference to efficiency estimates, which underlines the importance of dealing with these issues. Our claim is that these issues are resolved within our empirical framework, under its assumptions.

REFERENCES

Abbas, T., Awan, M. S., Aslam, M. A. and Waqas, M. 2011. Analyzing the Efficiency Differences among Basic Health Units in Sargodha District. *Journal of Economics and Behavioral Studies,* 3(1): 42-50.

Amado, C. A. d. E. F. and Santos, S. P. d. 2009. Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy,* 91(1): 43-56.

Appleby, J., Humphries, R., Thompson, J. & Galea, A. 2013. *How is the Health and Social Care System Performing? Quarterly Monitoring Report.* London: The King's Fund. Available online: http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/quarterly-monitoring-report-kingsfund-jun13.pdf.

Baltagi, B. 2008. *Econometric Analysis of Panel Data*. Chichester: John Wiley & Sons.

Bernet, P. M., Rosko, M. D., Valdmanis, V. G., Pilyavsky, A. and Aaronson, W. E. 2008. Productivity efficiencies in Ukrainian polyclinics: lessons for health system transitions from differential responses to market changes. *Journal of Productivity Analysis,* 29(2): 103-111.

Beveridge, W. 1942. Social insurance and Allied Services. London: HMSO.

Black, N. 2012. Declining health-care productivity in England: the making of a myth. *The Lancet*, 379(9821): 1167-1169.

Blundell, R. and Stoker, T. M. 2005. Heterogeneity and Aggregation. *Journal of EconomicLiterature*, 43(2): 347-391. Bojke, C., Castelli, A., Street, A., Ward, P. and Laudicella, M. 2013. Regional variation in the productivity of the English National Health Service. *Health Economics*, 22(2): 194-211.

Buckell, J., Smith, A., Longo, R. and Holland, D. 2013a. Health inefficiency and unobservable heterogeneity: empirical evidence from pathology services in the UK National Health Service. *Working paper series WP13-07, Academic Unit of Health Economics, University of Leeds*.

Buckell, J., Jones, R., Holland, D. and Batstone, G. 2013b. Efficient thinking: Introducing econometric techniques in pathology. *The Bulletin of The Royal College of Pathologists,* 164: 241-243.

Castelli, A., Jacobs, R., Goddard, M. and Smith, P. C. 2013. Health, policy and geography: Insights from a multi-level modelling approach. *Social Science & Medicine*, 92: 61-73.

Crivelli, L., Filippini, M. and Lunati, D. 2002. Regulation, Ownership and Efficiency in the Swiss Nursing Home Industry. *International Journal of Health Care Finance and Economics*, 2(2): 79-97.

Colombi, R., Martini, G. and Vittadini, G. 2011. A Stochastic Frontier Model with short-run and long-run inefficiency random effects. Working paper, Department of Economics and Technology Management, University of Bergamo.

D'Amico, F. and Fernandez, J.-L. 2012. Measuring Inefficiency in Long-Term Care Commissioning: Evidence from English Local Authorities. *Applied Economic Perspectives and Policy*, 34(2): 275-299.
Daidone, S. & Street, A. 2013. How much should be paid for specialised treatment? *Social Science & Medicine* 84: 110-118.
Department of Health Resource Allocation Working Party. 1976. Sharing Resources for Health in England. London: HMSO.

Department of Health. 2006. Report of the Review of NHS Pathology Services in England Chaired by Lord Carter of Coles. London: Department of Health. Available online: http://www.pathologists.org.uk/publications-page/Carter%20Report-The%20Report.pdf
Department of Health. 2007. Combining Age Related & Additional Needs (CARAN) Report. London: Department of Health.

Department of Health. 2008. Report of the Second Phase of the Review of NHS Pathology Services in England Chaired by Lord Carter of Coles. London: Department of Health. Available at http://microtrainees.bham.ac.uk/lib/exe/fetch.php?media=review_report_final_proof08.pdf
Farsi, M., Filippini, M. and Kuenzle, M. 2005. Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes. *Applied Economics*, 37(18): 2127-2141.
Felder, S. and Tauchmann, H. 2013. Federal State Differentials in the Efficiency of Health Production in Germany: An Artifact of Spatial Dependence? *European Journal of Health Economics*, 14(1): 21-39.
FTSE. 2010. Guide to Calculation Methods for the UK Series of the FTSE Actuaries Share Indices [online]. Accessed 30/3/2011. Available from http://www.ftse.com/Indices/UK_Indices/Downloads/FTSE_UK_Index_Series_Index_Rules.pdf

Gai, R. Y., Zhou, C. C., Xu, L. Z., Zhu, M., Wang, X. Z., Li, S. X., Zheng, W. G., Song, P. P., Yang, X. L., Fang, L. Y., Zhen, Y. C. and Tang, W. 2010. Health resource allocation and productive efficiency of Chinese county hospitals: data from 1993 to 2005. *Bioscience Trends,* 4(5): 218-24.
Greene, W. 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics,* 13(10): 959-980.
Greene, W. 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2): 269-303.
Greene, W. 2012a. *LIMDEP version 10*. New York: Econometric Software Inc.
Greene, W. 2012b. *LIMDEP Econometric Modelling Guide*. New York: Econometric Software Inc.

Gutacker, N., Bojke, C., Daidone, S., Devlin, N., Parkin, D. & Street, A. 2013. Truly Inefficient or Providing Better Quality of Care? Analysing the Relationship between Risk-Adjusted Hospital Costs and Patients' Health Outcomes. *Health Economics* 22: 931-947.
Halkos, G. E. and Tzeremes, N. G. 2011. A conditional nonparametric analysis for measuring the efficiency of regional public healthcare delivery: An application to Greek prefectures. *Health Policy*, 103(1): 73-82.
Ham, C. 2013. Balancing budgets or protecting patient safety. *BMJ*, 347.
Health Select Committee Public Expenditure. Thirteenth Report of the Session 2010-2012. Available online: www.publications.parliament.uk/pa/cm201012/cmselect/cmhealth/1499/149902.htm
Holland, D. & Trigg, G. 2011. Clinical Biochemistry/Chemical Pathology Core Report. NATIONAL PATHOLOGY BENCHMARKING REVIEW 2011/2012. Keele University.
Hu, H.-H., Qi, Q. and Yang, C.-H. 2012. Analysis of hospital technical efficiency in China: Effect of health insurance reform. *China Economic Review*, 23(4): 865-877.

Jacobs, R., Street, A. and Smith, P. 2006. *Measuring Efficiency in Health Care*, Cambridge: Cambridge University Press.

Jayasuriya, R. and Wodon, Q. 2007. Efficiency in Improving Health and Education Outcomes: Provincial and State-Level Estimates for Argentina and Mexico. *Estudios Economicos*, 22(1): 57-97.

Kontodimopoulos, N., Papathanasiou, N., Flokou, A., Tountas, Y. and Niakas, D. 2011. The Impact of Non-Discretionary Factors on DEA and SFA Technical Efficiency Differences. *Journal of Medical Systems*, 35(5): 981-989.

Law, S. M., Haghiri, M. and Nolan, J. F. 2010. Evaluating the Relative Efficiency of Cancer Treatments in Nova Scotia Using the Patient as the Unit of Analysis. *Canadian Journal of Regional Science*, 33(3): 163-176.

Lenard, M. L. and Shimshak, D. G. 2009. Benchmarking nursing home performance at the state level. *Health Services Management Research*, 22(2): 51-61.

Marmot, M. 2010. Fair Society, Healthy Lives. London: The Marmot Review.

Mundlak, Y. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1): 69-85.

Murillo-Zamorano, L. and Petraglia, C. 2011. Technical efficiency in primary health care: does quality matter? *The European Journal of Health Economics*, 12(2): 115-125.

Mutter, R., Greene, W., Spector, W., Rosko, M. & Mukamel, D. 2013. Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes. *Journal of Productivity Analysis* **39**: 101-110.

National Audit Office. 2012. Progress in making NHS efficiency savings. Available online: www.nao.org.uk/publications/1213/nhs_efficiency_savings.aspx.
Ng, Y. C. 2011. The productive efficiency of Chinese hospitals. *China Economic Review*, 22(3): 428-439.
Olsen, K. R. & Street, A. 2008. The analysis of efficiency among a small number of organisations: How inferences can be improved by exploiting patient-level data. *Health Economics* 17: 671-681.
Orea, L. and Kumbhakar, S. C. 2004. Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, 29(1): 169-183.
Pai, C.-W., Ozcan, Y. and Jiang, H. J. 2000. Regional Variation in Physician Practice Pattern: An Examination of Technical and Cost Efficiency for Treating Sinusitis. *Journal of Medical Systems*, 24(2): 103-117.

Pelone, F., Kringos, D. S., Valerio, L., Romaniello, A., Lazzari, A., Ricciardi, W. and de Belvis, A. G. 2012. The measurement of relative efficiency of general practice and the implications for policy makers. *Health Policy,* 107(2–3): 258-268.

Pesaran, M. H. and Tosetti, E. 2011. Large panels with common factors and spatial correlation. *Journal of Econometrics,* 161(2): 182-202.

Pilyavsky, A. and Staat, M. 2006. Health Care in the CIS Countries: The Case of Hospitals in Ukraine. *The European Journal of Health Economics,* 7(3): 189-195.

Pilyavsky, A. and Staat, M. 2008. Efficiency and Productivity Change in Ukrainian Health Care. *Journal of Productivity Analysis,* 29(2): 143-154.

Pitt, M. and L. Lee, .1981. The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry. *Journal of Development Economics* **9**: 43-64.
Purohit, B. C. 2008. Efficiency of the Health Care System: A Sub-state Level Analysis for West Bengal (India). *Review of Urban and Regional Development Studies,* 20(3): 212-225.

Purohit, B. C. 2010. Efficiency of health care system at the sub-state level in Madhya Pradesh, India. *Social Work Public Health,* 25(1): 42-58.

Rajitkanok, P. and Rosenman, R. 2008. Efficiency of Thai provincial public hospitals during the introduction of universal health coverage using capitation. *Health Care Management Science,* 11(4): 319-38.

Ramanathan, T. V., Chandra, K. S. and Thupeng, W. M. 2003. A Comparison of the Technical Efficiencies of Health Districts and Hospitals in Botswana. *Development Southern Africa,* 20(2): 307-320.

Scheffler, R. M., Wallace, N. T., Teh-Wei, H., Garrett, A. B. and Bloom, J. R. 2000. The impact of risk shifting and contracting on mental health service costs in California. *Inquiry - Blue Cross and Blue Shield Association,* 37(2): 121-33.

Schmidt, P. and Sickles, R. C. 1984. Production Frontiers and Panel Data. *Journal of Business & Economic Statistics,* 2(4): 367-374.

Sikka, V., Luke, R. D. and Ozcan, Y. A. 2009. The efficiency of hospital-based clusters: Evaluating system performance using data envelopment analysis. *Health Care Management Review,* 34(3): 251-261.

Smith, A. and Wheat, P. 2012. Estimation of cost inefficiency in panel data models with firm specific and sub-company specific effects. *Journal of Productivity Analysis,* 37(1): 27-40.

Smith, P., Mossialos, E. and Papanicolas, I. 2012. Performance measurementfor health system improvement: experiences, challenges and prospects. In: Figueras, J. and McKee, M., eds., *Health systems, health, wealth and societal well-being: assessing the case for investing in health systems.*, New York: McGraw-Hill, 247-280.

Smith, P. C. and Street, A. 2005. Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 168(2), 401-417.

Sørensen, T. H., Olsen, K. R. and Gyrd-Hansen, D. 2009. Differences in general practice initiated expenditures across Danish local health authorities—A multilevel analysis. *Health Policy,* 92(1): 35-42.

Theil, H. 1954. *Linear Aggregation of Economic Relations.* Amsterdam: North-Holland.

Valdez, S. and Molyneux, P. 2013. *An Introduction to Global Financial Markets*, 7th ed., Basingstoke: Palgrave Macmillan.

Veillard, J., Garcia-Armesta, S., Kadandale, S. and Klazinga, N. 2009. Internatonal health system comparisons: from measurement challenge to management tool' in

Smith, P., Mossialos, E., Papanicolas, I. and Leatherman, S., eds., *Performance Measurement for Health System Improvement*, Cambridge: Cambridge University Press.

Varela, P. S., Martins, G. d. A. and Favero, L. P. L. 2010. Production Efficiency and Financing of Public Health: An Analysis of Small Municipalities in the State of Sao Paulo—Brazil. *Health Care Management Science,* 13(2): 112-123.

Zafra-Gomez, J. L. and Muniz Perez, M. A. 2010. Overcoming Cost-Inefficiencies within Small Municipalities: Improve Financial Condition or Reduce the Quality of Public Services? *Environment and Planning C: Government and Policy,* 28(4): 609-629.

Zhang, N., Hu, A. and Zheng, J. 2007. Using Data Envelopment Analysis Approach to Estimate the Health Production Efficiencies in China. *Frontiers of Economics in China,* 2(1): 1-23.

Zhang, X., Hauck, K. and Zhao, X. 2013. Patient safety in hospitals – a Bayesian analysis of unobservable hospital and speciality level risk factors. *Health Economics,* 22(9): 1158-1174.