UNIVERSITY OF LEEDS

This is a repository copy of *Efficiency, heterogeneity and cost function analysis: empirical evidence from pathology services in the National Health Service in England*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83696/

Version: Accepted Version

Article:
Buckell, J, Smith, ASJ, Longo, R et al. (1 more author) (2015) Efficiency, heterogeneity and cost function analysis: empirical evidence from pathology services in the National Health Service in England. Applied Economics, 47 (31). pp. 3311-3331. ISSN 0003-6846

https://doi.org/10.1080/00036846.2015.1013617

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Efficiency, heterogeneity and cost function analysis: empirical evidence from pathology services in the National Health Service in England

J. Buckell[a,*], A. Smith[b,c], R. Longo[a] and D. Holland[d]

* Corresponding Author. 2.07 Charles Thackrah Building, Clarendon Road, Leeds, LS2 9LJ
E-mail: ts10jasb@leeds.ac.uk

[a]Academic Unit of Health Economics, University of Leeds, Leeds LS2 9LJ, UK
[b]Institute for Transport Studies, University of Leeds, Leeds LS2 9LJ, UK
[c]Leeds University Business School, University of Leeds, Leeds LS2 9LJ, UK
[d]Keele Benchmarking Service, Keele University, Keele ST5 5BG, UK

Keywords: Stochastic Frontier Analysis; Costs; Total Factor Productivity; Pathology

JEL classification C23 C81 D24 I11

Pathology services are increasingly recognised as key to effective healthcare delivery - underpinning diagnosis, long-term disease management and research. To the extent that pathology services affect a patient's treatment pathway, significant healthcare costs are influenced directly by the performance of these services. Given pressures on the UK Department of Health to make efficiency savings and that little is known about the efficiency of pathology laboratories, this area offers unlocked potential for efficiency gains. We adopt a time varying inefficiency model, with laboratory-specific time paths for inefficiency, to identify potential savings in pathology services based on a panel of 57 English laboratories over a five year period. We apply a range of approaches to account for observable and unobservable heterogeneity between laboratories. We find potential efficiency savings of 13% in pathology services in this sample, which implies the potential for an annual saving of £390m in pathology across the NHS. Our study also provides valuable insights into the impact of a range of factors influencing laboratory costs.

I. Introduction

The global financial crisis has increased pressure on public sector expenditure and so on the costs of the healthcare system in the UK. In response to this, the Nicholson Challenge has set out targets for efficiency savings of £20bn by 2015 in the UK National Health Service (NHS) (Health Select Committee, 2010). Financial pressure is expected to extend beyond 2015, with a funding gap of £30bn expected by 2020-21 (NHS, 2013). Thus, ensuring efficiency in all areas of healthcare is key.

There is a body of literature of both academic and other studies (e.g. think tanks such as the King's Fund, see Appleby et al., 2013) that has sought to measure inefficiency in the NHS. These may be at the macro or micro level. Typically, efficiency is measured by stochastic frontiers (SFs), Data Envelopment Analysis (DEA) or multivariate, multilevel modelling (MVML) in the academic literature, and using indicator analysis (such as mortality rates) in the non-academic literature.

At the macro level, the NHS itself is the unit of analysis, and is thus compared to other national healthcare services across the world. In Spinks and Hollingsworth (2009), the UK compared unfavourably (in terms of efficiency) amongst its OECD peers. However, the authors note that theoretical issues limit the interpretation of DEA results. Elsewhere, Smith and Street (2006) argue against the use of SFs at the macro level on theoretical grounds. Greene (2010) takes the view that using microeconomic tools at the macroeconomic level may be inappropriate. Practically, the usefulness of macro efficiency studies is somewhat restricted in the context of this policy challenge because these studies do not indicate where specific savings can be made within the NHS.

At the micro level, hospital studies dominate the national and international literature (Hollingsworth and Parkin, 1999; Jacobs et al., 2006; Hollingsworth, 2003; 2008). Secondary care is the largest tranche of NHS expenditure by far, totalling over £66bn in 2011-2012 (compared to the next largest, primary care, at £21bn) so significant savings potential is likely to reside here; at the same time the wealth of data available means that this is an area already well analysed in the more recent NHS-based literature (Farrar et al., 2009; Laudicella et al., 2010; Cooper et al., 2012; Gutacker et al., 2013; Siciliani et al., 2013; Daidone and Street, 2013). There is work in other areas of service delivery, primary care services for example (Szczepura, 1993; Giuffrida and Gravelle, 2001), however, because the outputs of these services are difficult to define and to measure, eliciting meaningful efficiency scores is challenging (Rosenman and Friesner, 2004; Lester and Roland, 2009; Amado and Santos, 2009; Murrillo-Zamorano and Petraglia, 2011; Longo et al., 2012). Perhaps it is unsurprising, then, that Hollingsworth (2008) finds no recent NHS primary care efficiency studies. The story is similar for other micro level services such as intermediate care.

Although there is a wide literature assessing efficiency performance of the NHS, new research is required since further gains are needed to meet the Nicholson Challenge. It has been argued that 'easy' efficiency savings have now been made across the NHS (National Audit Office, 2012). Further, surveys of NHS finance directors reveal growing scepticism about whether the Nicholson Challenge will be met at all (Appleby et al., 2013). Indeed, there is concern that financial pressure will continue beyond 2015 (Roberts et al., 2012). We therefore see potential in analysing diagnostic services which support healthcare delivery as an unturned rock to find new efficiency gains to contribute to the top-level policy goal. Specifically, we focus on pathology.

Pathology services account for an estimated 3-5% of the overall NHS budget, costing an estimated £2.5bn in 2005 (Department of Health, 2006). Although relatively small as a proportion of total health care spend, potential efficiency gains in these services are not limited to pathology itself. Pathology activity supports many front-line services and so savings in pathology services promote further gains elsewhere in the healthcare system (Veronesi et al., 1997; Buckell et al., 2013). The Carter Review (Department of Health, 2006) estimates 70-80% of all clinical decisions are affected by pathology analyses; thus good pathology practice can lead to cost savings along a patient's treatment pathway (Department of Health, 2006). There is evidence of unnecessary repeat testing (Department of Health, 2006), suggesting that inefficient practice is present in these services. Lastly, there is variation in the uptake of lean practice initiatives[1] meaning that there is likely variation in the magnitudes of efficiency in these services. Therefore, there are likely significant gains to be made by encouraging best practice in pathology services to contribute to the policy objective of achieving efficiency savings. This study aims specifically to identify the level of inefficiency in pathology services in order to measure the extent of savings possible in this area.

The current approach to measuring inefficiency in pathology in the NHS is performance indicator analysis (such as cost per test carried out); (Healthcare Commission, 2007; Department of Health, 2008; Liebmann, 2011; Holland et al., 2012). These are partial measures which do not fully reflect all the factors affecting the costs of provision under different circumstances (for example, scale properties or sources of operational heterogeneity between providers). This point has been established in the wider health context (Goddard and Jacobs, 2009; Street et al., 2011). We use the data collected and analysed by the Keele University Benchmarking Unit (Holland et al., 2012), but extend the analysis by utilising an

[1] NHS Institute for Innovation and Improvement: Pathology lean practice case studies, http://www.institute.nhs.uk/quality_and_value/lean_thinking/leean_case_studies.html

econometric framework to give a single measure that captures the overall efficiency of pathology services. Our model takes account of a range of factors influencing costs, whilst controlling for unobservable heterogeneity.

We use SFs which have been applied widely in health at the micro level (Street, 2003; Farsi and Filippini, 2005, 2008; Herr, 2008; Hollingsworth, 2008; Olsen and Street, 2008; Rosko & Mutter, 2008; Sorensen et al., 2009; Herr et al., 2011). We adopt a particular SF method with attractive properties in respect of analysing efficiency change over time; this method has been applied by economic regulators outside health for that reason (Smith, 2012). To our knowledge, no SF (or other efficiency measurement tool such as DEA) work has been conducted on pathology laboratories, meaning that our application is the first of its kind[2].

The remainder of this paper is set out as follows. In section 2 we review the existing literature on pathology performance and argue that an econometric framework can offer a useful extension to these analyses. In section III we set out our methods, data and empirical specification. In section IV we give our results, in section V we discuss our results, and in section VI we conclude.

II. Performance Measurement in pathology

Pathology services are increasingly recognised as key support for a range of services across the NHS. As demand for NHS services increases in general, demand for pathology services increases (as derived demand). Faced with increasing demand and falling income (Department of Health, 2006), the performance of pathology services is coming under ever-

[2] If pathology is classed as diagnostic medicine, then there exists some SF work in this area (Dismuke & Sena, 1999). However, this study concerns patient-based, in-hospital activity such as computerised axial tomography (CAT) scans, whereas our study involves pathology laboratories – which are independent of their host hospitals and do not have direct patient contact – conducting blood and tissue tests. We therefore view pathology services as distinct from this kind of diagnostic medicine.

increasing scrutiny. Therefore, rigorously measuring the performance of laboratories is critical. Typically, pathology laboratories are situated within NHS trusts (see below).

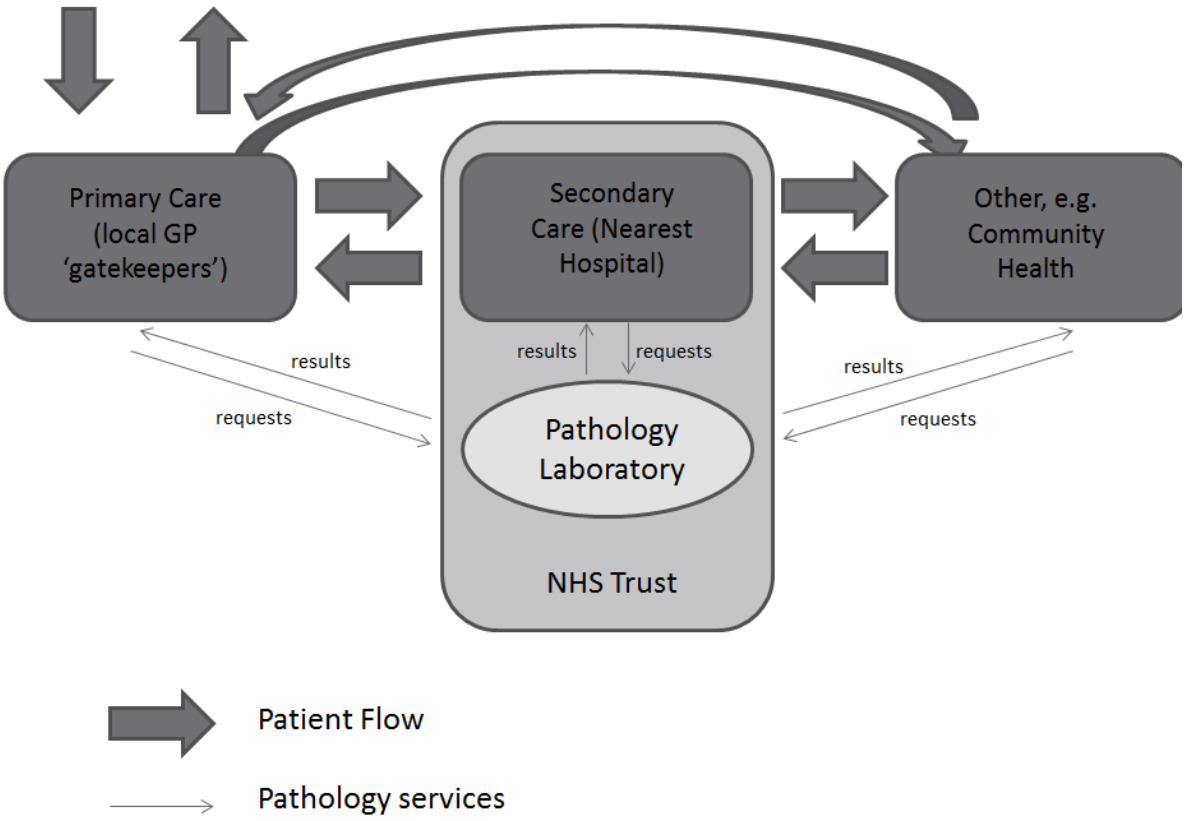

Fig. 1: Schematic of pathology services

As can be seen from Fig. 1, as patients move around the healthcare system, diagnostic services are requested and performed. As activity occurs, information is recorded and used for analysis of these services.

Major reviews of NHS pathology services include the Carter Report (Department of Health, 2006), and the associated follow up report which included pilot studies of services (Department of Health, 2008); the Healthcare Commission's study (2007); the NHS confederation (2010); and the Keele University Benchmarking project (Holland et al., 2010;

2011; 2012)[3]. There is a growing body of evidence on these services, and good quality data available; a summary of these studies' analyses is provided in Table 1.

[3] Some key performance indicators are being introduced, but have not yet been employed (Liebmann, 2011).

Study	Year	Number of Sites	Type of study	Summary of Key Points
Department of Health	2006	163	Qualitative	Full qualitative analysis of pathology services. Identified key areas for performance improvement - workforce balance, economies of scale, information systems adoption, out of hours working, network activity. Recommended pilot studies conducted. Noted that geographical location may be a source of cost heterogeneity.
Healthcare Commission	2007	163	Quantitative	Breakdown by pathology discipline comparative cost per test analysis; requests:staff and tests:staff ratios used; descriptive statistics for out of hours operation, information systems adoption, use of automated services, network activity; recognised that tests for primary care may be cheaper than for secondary care; noted the issue of tests:requests as a potential source of performance variation. Foundation trusts may take a commercial approach to service provision.
Department of Health	2008	12	Quantitative	Breakdown by pathology discipline (e.g. biochemistry) comparative cost analysis; some economies of scale observation; little control for heterogeneity; savings estimate £250m (extrapolated results nationally from 12 pilot studies).
NHS Confederation	2010	163	Qualitative	Identifies variation in practice; difficulty in monitoring staff leads to variation in practice; workforce balance, IT systems adoption, leadership and network activity as key areas for performance improvement.
Keele Benchmarking	2012	84	Quantitative	Breakdown by pathology discipline (e.g. biochemistry, hystocytology); test volumes descriptive statistics; productivity indicators; 5 year trend analysis of outputs and productivity indicators; expenditure of laboratories; quality indicators (e.g. turnaround times)

Table 1: Pathology literature

Table 1 describes the outcomes of each of the studies. The quantitative analyses above use performance indicators to judge the performance of NHS pathology laboratories (e.g. cost per test ratios, staff per test, turnaround times, test to request ratios). The use of these indicators is widespread in NHS pathology and across the world (Valenstein et al., 2001; Kiechle and Main, 2002; Price, 2005; France and Francis, 2005), but there are limits to their ability to reflect the entire operation of a laboratory. Moreover, in health markets, indicators can be targeted for gaming (Propper and Wilson, 2003; Propper et al., 2008; Mutter et al., 2008; Palangkaraya and Yong, 2013), or relying solely on indicators can lead to unintended consequences (Bird et al., 2004; Cots et al., 2011). Lastly, judging a single unit's performance across several indicators may be difficult if the values conflict.

An econometric framework is proposed to overcome these issues. Our measure of cost efficiency yields a single efficiency score capturing overall performance which is easily interpreted (bounded by zero and one). Gaming is no longer an issue since the entire production process is modelled[4] .

A further key advantage of the econometric approach is that it is underpinned by economic theory and stochastic frontier analysis is used widely across many sectors, including health (Kumbhakar and Lovell, 2000; Hollingsworth, 2008). In addition, we can analyse the temporal pattern of laboratory inefficiency, which NHS staff have indicated as a desirable feature of performance analysis (Hollingsworth and Peacock, 2008). Finally, econometric analysis allows us to value the impact of some of the issues noted in the qualitative studies (Table 1), such as the ratio of primary care tests on costs – as raised in the Healthcare Commission study (2007), which is useful information in the policy context.

[4] We use operating costs rather than total costs (including capital charges), meaning the production process is not strictly entirely modelled. Capital costs are budgeted centrally at trust (hospital) level, rather than laboratory level, meaning assigning specific capital charges to laboratories can only be estimated. We note that this has been found in pathology elsewhere, e.g. New Zealand (France and Francis, 2005). Moreover, this is not particular to pathology (Drummond et al., 2005, pp. 64).

III. Methods

Stochastic frontiers (Aigner et al., 1977; Meeusen and van Den Broeck, 1977) are econometric tools used to estimate the level of inefficiency of firms or decision making units (DMU) in a sample. Laboratory costs are our metric of interest. Our economic SF model for pathology, derived from a basic cost function, takes the form,

$$c = f(y, w, z, q, t) + u + v \quad (1)$$

Where c are costs, y represents output, w represents input prices, z represents the observable heterogeneity, q represents quality and t represents time. As standard for SFs, u represents the inefficiency and v represents random statistical noise.

As standard in the literature, output and input prices are considered exogenous, which is obvious for input prices and reasonable for output levels given that the laboratories do not choose their level of output. In the case of pathology, using the work of previous studies (see table 1), the operational characteristics of the pathology operating environment can be identified and variables are used to capture these where data are available (the z vector). Otherwise, methods for capturing unobservable heterogeneity are employed.

For service quality, although measures of quality in pathology services are not as complex as in the treatment of patients (Smith and Street (2013) discuss the multi-dimensional nature of patient treatment quality), this remains an issue for our study. Each of the laboratories in our sample has acquired quality accreditation[5]. Our understanding of accreditation is that it represents a baseline level of quality. Therefore, we recognise that there may well be laboratory-specific variation in quality over and above this baseline level. This is one reason for which we apply empirical controls for unobserved heterogeneity; that is, quality that is

[5] Clinical Pathology Accreditation: http://www.cpa-uk.co.uk/

not captured in the accreditation is absorbed into the control for unobserved heterogeneity rather than absorbed by the inefficiency component of the model.

A set of five models SF is used to model inefficiency. These include a generalised least squares random effects model[6], see Kumbhakar and Lovell (2000). We refer to this as REM. We use a Pitt and Lee (1981) stochastic frontier with time invariant inefficiency, which we refer to as P&L. Next, we use a Battesse and Coelli (1992) SF with time varying inefficiency. We refer to this as BC92. Our penultimate model is that of Cuesta (2000), which is a SF with firm-specific (or in our case, lab-specific) time-varying inefficiency. We refer to this as Cuesta. Finally, we use a true random effects model (Greene, 2005). We refer to this as TRE. See table 2 for econometric specification.

The REM is used to give 'baseline' values for both parameter estimates and for inefficiency (using the GLS procedure outlined in Kumbhakar and Lovell (2000)). Parameter estimates from these models do not rely on the distributional assumptions of the SFs[7] and so parameter estimates are used to validate those derived from the frontiers.

The P&L model assumes time-invariant inefficiency. The BC92 fits a time trend to the inefficiency - the η parameter (table 2) - which subjects all firms' efficiency scores to a common direction of change over time. The Cuesta model is a generalisation of this, allowing estimation of independent firm efficiency time trends: individual ηs for each laboratory[8]. This means firms can 'catch up' relative to others over time and the efficiency rankings of the laboratories can change over time, which are realistic features. This point is particularly relevant in a policy context, and this model has been used by regulators in other sectors, e.g. rail (Smith, 2012). Alvarez et al. (2006) further note that a key advantage of this model is that

[6] Hausman tests (1978) consistently favoured RE over FE estimation; we are also interest in examining time-invariant variables which we are unable to do in a FE framework.
[7] Due to an unbalanced panel, a Baltagi & Li (1990) adaptation of the Breusch-Pagan (1980) test has been used and confirms the use of panel methods.
[8] Within this framework, the temporal pattern of inefficiency can be tested statistically, which is a key advantage over alternative approaches such as Cornwell et al. (1990).

it enables the unrealistic assumption of independence in inefficiency over time (a problem that plagues many comparator models) to be relaxed.

The TRE model claims to delineate efficiency from unobservable heterogeneity by including a time-invariant, firm-specific term in the model to capture unobserved factors, in addition to the inefficiency term (Greene, 2012a). A potential drawback of this model is that efficiency scores are independent over time, meaning that time trends of firms cannot be tested statistically. Additionally, this model assumes that all the time-invariant variation in the cost function that is not explained by the regressors is unit-specific heterogeneity and not inefficiency; this is not necessarily the case as some time invariant persistent inefficiency may also be present.

To these models, we test three alternative specifications to examine heterogeneity. First, a basic cost function with output, input prices and time is estimated. By including a time trend in the cost function, we separate exogenous change in costs over time from cost inefficiency (Kumbhakar and Lovell, 2000).

In the second, we add the vector, z, of observable heterogeneity variables. These include the number of primary care tests (which are thought to be less costly than other tests), and the test to request ratio which captures the variation in the number of tests per request, which varies between laboratories, and is therefore a source of heterogeneity. Another source is the geographical setting of the laboratory: metropolitan, urban or rural (following Department of Health, 2006, see table 1). This will be referred to as the TYPE of laboratory. It has been suggested that pathology demands of inner city laboratories are much different to those in rural areas. Further, the foundation status[9] of a trust is seen to motivate it to act more commercially (Healthcare Commission, 2007, see table 1; Marini et al., 2008), which is

[9] Foundation status of a NHS trust (a trust is a hospital or small group of hospitals) means that it operates under an independent, not-for-profit regime, allowing it financial autonomy which it does not have without having foundation status (Marini et al., 2008). Trusts apply for foundation status, which is granted by the regulator, monitor, if the trust has satisfied the regulator of its financial competence. Foundation status has not been awarded to all NHS trusts.

expected to be extended to their pathology services. Lastly, data are available on whether the laboratories provide teaching services.

The third specification finally adds dummy variables to capture unobservable heterogeneity (e.g. IT infrastructure/maturity, network activity) (Arocena et al., 2012). We use the strategic health authority dummy variables and then group them by region for parsimony.

We refer to the specifications as s(i), s(ii) and s(iii).

Finally, after having used this testing process to select a model, we exploit the fact that the SF framework is based on a cost function to examine the cost elasticity properties across the output range and derive average and marginal costs in pathology production (AC and MC hereafter). We note that this is a key advantage of this method over DEA as an alternative. Focus is given to this aspect of production because this is a popular theme of interest throughout the literature (table 1), because there is little empirical evidence on this issue, and because of the growing membership of laboratories to local networks, which is encouraging the pooling of output); see Department of Health (2011).

Empirical Specification

First, for functional form, we test between a Cobb-Douglas and a translog specification to approximate our economic model in (1). A translog nests a Cobb-Douglas and we can readily test down. A translog has some appealing empirical and economic features: its flexible nature means it provides a second-order differential approximation to any unknown function $f(.)$ (as in Equation (1)) (Kumbhakar and Hjalmarsson, 1995); it does not impose restrictions on substitution possibilities; and allows economies of scale to vary with output levels (Christensen and Greene, 1976).

Logarithms are taken to give Farrell (1957)-type radial measures of inefficiency[10]. The translog representation is estimated for each model,

$$
\begin{aligned}
&ln\, c_{it} \\
&= \alpha_0 + \beta_1\, ln\, y_{it} + \frac{1}{2}\beta_2\, (ln\, y_{it})^2 + \beta_3\, ln\, wl_{it} \;\; + \frac{1}{2}\beta_4\, (ln\, wl_{it})^2 \\
&+ \sum_{a=1}^{2} \beta_a lnz_{it} + \frac{1}{2}\sum_{a=1}^{2} \beta_a\, (ln\, z_{it})^2 + \beta_5 lny_{it}.\, lnwl_{it} + \sum_{n=1}^{1}\sum_{a=1}^{2} \beta_{na}\, lny_{it}.\, ln\, z_{it} \\
&+ \sum_{b=1}^{1}\sum_{a=1}^{2} \beta_{ba} ln\, wl_{it}.\, ln\, z_{it} + \beta_6 lnz_{1t}.\, lnz_{2t} + \sum_{c=1}^{4} \beta_c z_i + \sum_{d=1}^{3} \beta_d \omega_r + \beta_7 t \\
&+ \varepsilon_{it}
\end{aligned}
\qquad (2)
$$

Where c_{it} are operating costs; y_{it} is output; wl_{it} are labour input prices; z_{it} are exogenous variables including tests for primary care and the test to request ratio; z_i are laboratory-specific, time-invariant dummy variables for the following: foundation status, teaching status and laboratory type[11]; ω_r are regional dummy variables to capture unobservable heterogeneity; and t is a time trend capturing real cost changes over time (in this sample). Then, ε_{it} is decomposed into u_{it} and v_{it} which are inefficiency and statistical noise, respectively (see table 2 below for detailed specifications of each model).

To decide on a preferred model, a number of statistical tests are applied[12]. We test functional form using a Wald test[13].

Next, we test between the three specifications from above, by which we mean either no heterogeneity variables s(i); observable heterogeneity variables only s(ii); and observable and

[10] Variables are mean scaled to allow direct interpretation of the first order terms.
[11] Types of laboratory include rural, urban and metropolitan; rural is the reference case for modelling.
[12] Lai and Huang (2010), pp. 3, lament that "there are only limited systematic treatments of tests or model selection criteria in the existing SF literatures."
[13] H_0: additional translog terms (squared and cross terms) are jointly equal to zero.

unobservable heterogeneity variables[14] s(iii). We use LR tests for this. We refer to this as TEST 1.

We then test between each efficiency model, by which we mean one of the 5 different efficiency models (REM, P&L, BC92, Cuesta, TRE), using a LR test[15] for nested models (which we refer to as TEST 2) and a Vuong test (1989) for non-nested models[16] (which we refer to as TEST 3).

In total, there are 30 models to be estimated[17]. 15 models are reported for comparison which represents our full set of models once the test for functional form has been applied. LIMDEP software (Greene, 2012b) is used for estimation.

Efficiency models

Table 2 below shows the econometric specifications of our range of models estimated.

	REM	P&L	BC92	CUESTA	TRE
Firm-specific component, α_i	$iid(0,\sigma_\alpha^2)$	$iid(0,\sigma_\alpha^2)$	$iid(0,\sigma_\alpha^2)$	$iid(0,\sigma_\alpha^2)$	$N(0,\sigma_\alpha^2)$
Random Error, ε_i	$iid(0,\sigma_\varepsilon^2)$	$\varepsilon_{it}=u_{it}+v_{it}$ $u_{it}\sim\lvert N(0,\sigma_u^2)\rvert$ $v_{it}\sim N(0,\sigma_v^2)$	$\varepsilon_{it}=u_{it}+v_{it}$ $u_{it}\sim\lvert N(0,\sigma_u^2)\rvert$ $v_{it}\sim N(0,\sigma_v^2)$	$\varepsilon_{it}=u_{it}+v_{it}$ $u_{it}\sim\lvert N(0,\sigma_u^2)\rvert$ $v_{it}\sim N(0,\sigma_v^2)$	$\varepsilon_{it}=u_{it}+v_{it}$ $u_{it}\sim\lvert N(0,\sigma_u^2)\rvert$ $v_{it}\sim N(0,\sigma_v^2)$
Inefficiency	$\hat{\alpha}_i-min\{\hat{\alpha}_i\}$	$E[u_{it}\lvert u_{it}+v_{it}]$	$E[u_{it}\lvert u_{it}+v_{it}]$	$E[u_{it}\lvert u_{it}+v_{it}]$	$E[u_{it}\lvert\alpha_i+\varepsilon_{it}]$
Time Trend			$u_{it}=\exp[\eta(t-T)].u_i$	$u_{it}=\exp[\eta_i(t-T)].u_i$	

Table 2: Econometric Specifications of models

Merging laboratories

A feature of recent pathology services is that, following recommendations from the Carter Review, laboratories in close proximity are increasingly beginning to pool their production

[14] H_0: observable or unobservable heterogeneity variables are jointly equal to zero.
[15] H_0: log likelihood model (a) is equal to log likelihood model (b)
[16] H_0: model (a) is equal to model (b)
[17] 2 (functional forms) x 3 (heterogeneity variable specifications) x 5 (types of efficiency model)

(Department of Health, 2006; 2009). A natural question arises as to what happens to the costs of production when laboratories merge. This is, of course, tied closely to the issue of economies of scale, which is of great interest to NHS policy makers and policy makers more widely.

In our data, there are no examples of laboratory mergers. However, it is possible to use the model to simulate the effects of laboratories merging to shed some light on this issue: we can simply compare the sum of the predicted merged laboratory costs and the sum of the predicted unmerged laboratory costs. We do this for laboratories in the final year of the dataset.

To operationalise the merged scenario, we merge the smaller laboratories with each other. We define a "small laboratory" as one whose output (number of requests) is lower than the sample median. We then merge the largest "small laboratory" with the smallest "small laboratory", the second largest with the second smallest, and so on. We assume the larger laboratory absorbs the smaller; we thus assume the characteristics (i.e. foundation status, teaching status, region, etc.) of the larger laboratory for computing merged cost estimates. We are interested in the proportional change in total costs that would occur if small laboratories were to merge, thus we compute the following ratio,

$$\frac{\sum_{i=1}^{I} E\left(c_{i,T}\middle|{x_{it}}'\beta\right) - \sum_{j=1}^{J} E\left(c_{j,T}\middle|x_{it}'\beta, y > \tilde{y}\right)}{\sum_{i=1}^{I} E\left(c_{i,T}\middle|{x_{it}}'\beta\right)} \tag{3}$$

where $E\left(c_{i,T}\middle|{x_{it}}'\beta\right)$ is the conditional expectation of costs for laboratory i in its final year, T. The ${x_{it}}'\beta$ is the estimated cost function, y is output and $y > \tilde{y}$ denotes all output is greater than the (original) sample median, that is, laboratories with output lower than the median have merged. $\sum_{i=1}^{I} E\left(c_{i,T}\middle|{x_{it}}'\beta\right)$ is the sum of the predicted costs across all unmerged laboratories and $\sum_{j=1}^{J} E\left(c_{j,T}\middle|x_{it}'\beta, y > \tilde{y}\right)$ is the sum of predicted costs across all merged

laboratories. As a result of simulation, of the full sample of 57 laboratories, 28 "small" laboratories are merged into 14, thus reducing the number of laboratories from 57 to 43. Therefore, I, the number of unmerged laboratories, is 57 and J, the number of merged laboratories, is 43.

Given the specification of our model (see equation (2)), there is an issue around retransformation of logged (predicted) costs (Manning, 1998). When the disturbance of the error term is normal, $\hat{\varepsilon} \sim N(0, \sigma^2(x))$, then a straightforward correction can be made,

$$E\left(c_{i,T} \middle| x_{it}'\beta\right) = e^{x_{it}'\beta + 0.5\sigma^2(x)} \tag{4}$$

where the uncorrected estimate is an underestimate since,

$$e^{x_{it}'\beta + 0.5\sigma^2(x)} > e^{x_{it}'\beta} \tag{5}$$

However, normality is an invalid assumption in our case as the stochastic frontier model does not, by definition, assume a normally distributed disturbance. Thus, an approach is required that can account for non-normally distributed errors. Therefore, as suggested by Greene (2012c, pp. 123), we use the smearing estimator proposed by Duan (1983). Thus our predictions of laboratory costs are,

$$E\left(c_{i,T} \middle| x_{it}'\beta\right) = h^0 e^{x_{it}'\beta} \tag{6}$$

where,

$$h^0 = \frac{1}{n}\sum_{i=1}^{I} e^{\hat{\varepsilon}_i} \tag{7}$$

where n denotes the number of observations and $\hat{\varepsilon}_i$ are the fitted residuals.

Data

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[18] (187 observations); accordingly we use maximum likelihood estimation (Baltagi, 2008) (except the REM which uses GLS and the TRE which uses simulated maximum likelihood). The sample represents approximately one third of the 163 NHS pathology laboratories in England. From table 3, 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. There is an almost even spread of laboratories amongst strategic health authorities (and therefore across England).

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 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 (for which two measures are available: the number of tests and the number of requests), input prices of labour (from the UK labour force survey) and exogenous variables including the number of tests for general practice (primary care) and dummy variables for the foundation status of the host trust, for the pathology service providing teaching, for the laboratory type (metropolitan, urban, rural) and for the strategic health authority in which the pathology service is located. Service quality is assumed given that laboratories have been accredited as noted earlier.

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: because the labour force survey data is collected and constructed independently from our study data, which would not be the case using the

[18] In our sample, 27 laboratories are observed twice, 7 are observed 3 times, 2 are observed 4 times and 21 are observed in every year – 5 times.

NHS-based data[19]. In addition, this data is a reflection of the true labour market conditions, which is not necessarily the case with the NHS data. Lastly, the NHS equivalent data is constructed using staff numbers which implies the measure may be correlated with output, which may lead to undesirable statistical issues such as collinearity. 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[20]. Strategic health authorities are, following initial modelling, combined to form regional dummy variables for London, the South, the Midlands and the North using a Wald test procedure (Greene, 2012a).

One available measure of clinical quality was available for analysis: turnaround times of tests. We did not use this for three major reasons. First, as an indicator, this is an incomplete measure of clinical quality (i.e. there are other dimensions of quality which may vary). This may induce measurement error if used to capture quality in our cost function. Second, some laboratories, although recording turnaround times, do not make efforts to reach targets as they are not enforced. This means that this measure is likely to give a skewed reflection of this (partial) measure of quality. Third, the data completeness and validity is much lower than for the remainder of collected data (partly as some labs do not pay a great deal of attention to turnaround times).

[19] Mutter et al. (2013) demonstrate using healthcare data that endogeneity can bias efficiency scores.
[20] As this variable is constructed using a variable that is also in the models, we check the correlation of the two variables for collinearity concerns. The correlation between the two variables is -0.34. We therefore do not see this as an issue. In any case, we note that collinearity is less an issue in panel data models than in cross sectional or time series alternatives (Baltagi, 2008).

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
TYPE: Metropolitan	Dummy	(0 or 1)		
TYPE: Urban	Dummy	(0 or 1)		
TYPE: Rural	Dummy	(0 or 1)		
Foundation Trusts	Dummy	(0 or 1)		
Teaching Laboratories	Dummy	(0 or 1)		
REGION: London	Dummy	(0 or 1)		
REGION: South	Dummy	(0 or 1)		
REGION: Midlands	Dummy	(0 or 1)		
REGION: North	Dummy	(0 or 1)		

Table 3: Descriptive Statistics

IV. Results

Cost function parameters

Across the range of models estimated (table 4), a number of general observations can be made. Cost elasticity with respect to output implies economies of scale (which we refer to as size – see later section) in pathology production (the first order parameters are elasticities at the sample mean; we go on to explore how these vary with output later in this section). Real

operating costs appear to be decreasing over time as indicated by the negative coefficient on the time trend variable. Operating costs in pathology laboratories are higher for those which have high test to request ratios, are located in metropolitan and urban locations (relative to rural laboratories), provide teaching services and are in the Midlands (relative to the Northern laboratories). Operating costs are lower for foundation trust laboratories and for those located in London or the South (relative to the North). There was no clear finding as to the effect of GP tests on laboratory operating costs, where the effect appears negative in two models, positive in another and not statistically significant in any other.

Statistical testing and inefficiency model selection

Wald tests strongly and consistently favoured the translog functional form (the null being the Cobb-Douglas). Test 1 finds the s(ii) and s(iii) heterogeneity variables jointly significant additions to the models in all cases (table 5). Test 2 strongly favours the Cuesta model over the BC92 and P&L. Test 3 favours the Cuesta model over the TRE model[21]. Therefore our preferred inefficiency model is Cuesta s(iii) based on statistical criteria. Indeed, this model is preferred a priori because of how it deals with efficiency change over time (see section III for details). A significant lambda value (table 4) confirms the presence of inefficiency[22].

[21] We are aware that the Vuong test has no degrees of freedom restriction, meaning that it imposes no penalty for additional parameters estimated and so is likely to, in this case, favour the Cuesta model which has more parameters than the TRE model. Therefore, as a robustness check, we have also tested the P&L (which has fewer parameters than the TRE) against the TRE, and the test favours the P&L. Because our LR test preferred the Cuesta to the P&L, and the Vuong preferred the P&L to the TRE, we prefer the Cuesta to the TRE.

[22] In addition, we have tested the presence of inefficiency using the LR test procedure outlined in Coelli et al. (2005) pp.258, which also confirms our result, but we do not report the test results here.

Dependent Variable: Operating cost (OPEX)	Specification														
	s(i) c = y, w, t					s(ii) c = y, w, t, z – observable					s(iii) c = y, w, t, z - observable, z - unobservable				
	Model														
	REM	P+L	BC92	CUESTA	TRE	REM	P+L	BC92	CUESTA	TRE	REM	P+L	BC92	CUESTA	TRE
PARAMETER VALUES															
CONSTANT	6.55***	6.32***	6.31***	6.31***	6.60***	6.55***	6.40***	6.39***	6.39***	6.61	6.53***	6.42***	6.42***	6.35***	6.54***
OUTPUT	0.43***	0.29***	0.31***	0.30***	0.74***	0.67***	0.55***	0.62***	0.35***	0.99***	0.67***	0.58***	0.64***	0.44***	0.93***
INPUT PRICES	0.61***	0.52**	0.54**	0.68***	0.64***	0.64***	0.53***	0.49**	0.59***	0.84***	0.83***	0.80***	0.89***	1.30***	1.04***
YEAR	-0.01**	-0.01	-0.01	-0.01**	-0.02***	-0.01**	-0.01*	0.01	-0.01	-0.01	-0.01**	-0.01*	0.01	-0.01***	-0.01***
GP_TESTS						0.01	0.03	0.02	0.07*	-0.14***	0.01	0.02	0.03	0.06	-0.08***
TES:REQ						0.23***	0.19**	0.21***	0.02	0.47***	0.24***	0.21***	0.21***	0.12**	0.48***
TYPE: METROPOLITAN						0.13***	0.14***	0.14***	0.15***	0.09***	0.14***	0.15***	0.15***	0.16***	0.10***
TYPE: URBAN						0.03	0.04*	0.04	0.05*	0.01	0.02	0.02	0.01	0.02	0.01***
FOUNDATION						-0.06**	-0.07***	-0.07***	-0.11***	-0.06	-0.04	-0.06**	-0.06***	-0.07***	-0.04***
TEACHING						0.03	0.04*	0.03	0.01	0.01	0.04	0.05**	0.03	0.02	0.03***
REGION: LONDON											-0.02	-0.05	-0.07**	-0.16***	-0.02***
REGION: SOUTH											-0.03	-0.04	-0.05*	-0.01	-0.01***
REGION: MIDLANDS											0.08**	0.10***	0.09***	0.10***	0.08***
EFFICEINCY FIGURES															
mean	0.71	0.81	0.81	0.79	0.99	0.76	0.87	0.88	0.82	1.00	0.77	0.90	0.90	0.87	1.00
s.d.	0.10	0.11	0.11	0.12	0.00	0.08	0.08	0.08	0.11	0.00	0.07	0.07	0.07	0.10	0.00
lambda		5.11***	5.17***	13.01***	3974.52		3.15***	3.33***	11.97***	0.00		2.67***	3.04***	8.38***	552028
eta			-0.01					-0.07*					-0.11**		

Table 4: Estimation outputs

Notes: *,** and *** denote significance at the 10%, 5% and 1% level, respectively.

LR Statistic Tests for Heterogeneity Variables: TEST 1

Model		P&L	BC92	CUESTA	TRE
Restriction of S(ii) to S(i): Observable heterogeneity variables	(d.f.: 13,13,13,12)	44.6***	48.00***	44.82***	91.04***
Restriction of S(iii) to S(ii): Unobservable heterogeneity variables	(d.f.: 3,3,3,4)	14.86***	17.70***	8.38***	38.60***

LR Statistic Tests for Model Selection (nested models only): TEST 2

			CUESTA v. P&L	CUESTA v. BC92
Specification (i): Basic Cost function		(d.f.: 57, 56)	166.84***	166.70***
Specification (ii): Observable Heterogeneity		(d.f.: 57, 56)	167.00***	163.32***
Specification (iii): Regional Dummies for Unobserved Heterogeneity		(d.f.: 57, 56)	160.52***	154.00***

Vuong Test Statistic: TEST 3

TRE specification (iii) vs. Cuesta model specification (iii)	V = -9.066***

Model Log Likelihood Function Values and degrees of freedom (K)

Model	P&L	BC92	CUESTA	TRE
Specification (i): Basic Cost function	198.80	198.97	282.22	135.81
K	9	10	66	10
Specification (ii): Observable Heterogeneity	221.13	222.97	304.63	181.33
K	22	23	79	23
Specification (iii): Regional Dummies for Unobserved Heterogeneity	228.56	231.82	308.82	200.63
K	25	26	82	26

Table 5: LR specification and model selection

Notes: *,** and *** denote significance at the 10%, 5% and 1% level, respectively.

Inefficiency estimates

From table 4, the mean efficiency estimate from our preferred model is 0.87. On average, efficiency is computed as decreasing slightly amongst pathology laboratories over time (which is in agreement with the BC92 s(iii) model[23] in table 4, given their eta coefficients) from 0.87 in 2007 to 0.86 in 2011. Fig 2 shows the cost efficiency estimates of laboratories over time. The bar in Fig. 2 is at efficiency = 1, i.e. full efficiency. Groups of points correspond to each individual laboratory, e.g. observations 1-5 are the efficiency estimates for laboratory 1 in years 1 to 5, observations 6 to 10 are laboratory 2 in years 1-5, and so on. We do not find the problem of efficiency scores dropping off the frontier in the final year of the sample, which has been a concern for other applications of this model (Wheat and Smith, 2012). In addition, we find that many of the laboratory-specific etas are statistically significant. Those that were not tended to be the firms that are on the frontier (and thus have little or no inefficiency change over time), which can be seen in figure 2.

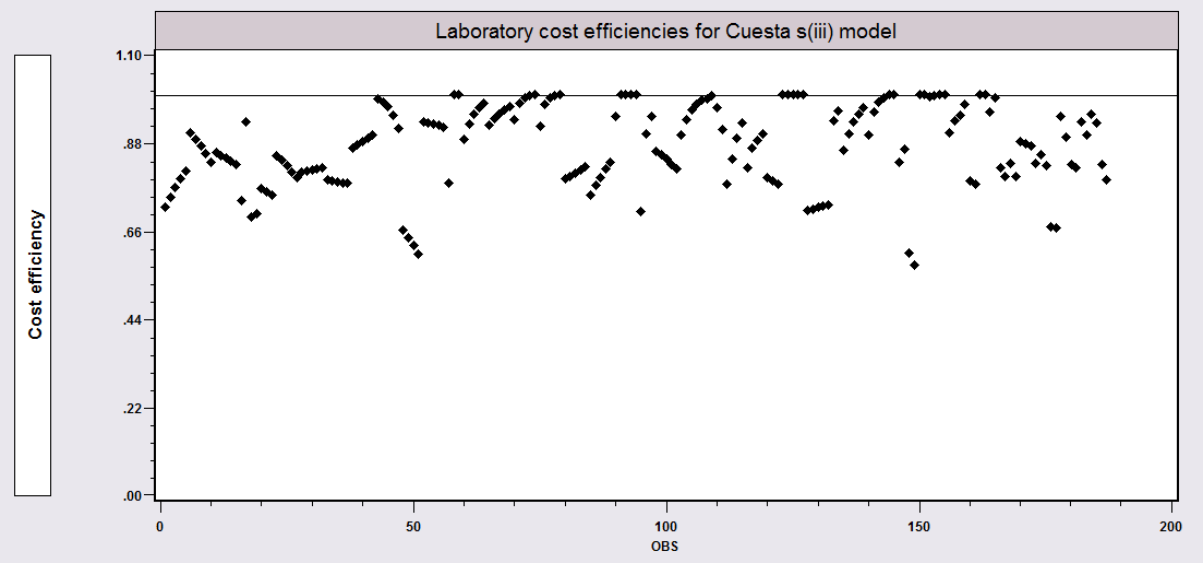

Fig. 2: Laboratory cost efficiency estimates over time

[23] Which is preferred of the three candidate BC92 models, see table 5

Elasticity of cost, average and marginal costs

Our set of models give estimates of the elasticity of cost with respect to output at the sample mean in the range of 0.29-1.04 (table 4) and is 0.44 in the preferred model. However, a more informative approach is to examine how this elasticity changes with the scale of the operation, proxied by output (Fig. 2), using our preferred model. Using this elasticity, we are able to further estimate AC and MC per request using fitted values from the model (see Wheat and Smith, 2008, for details) (Figs 4 and 5).

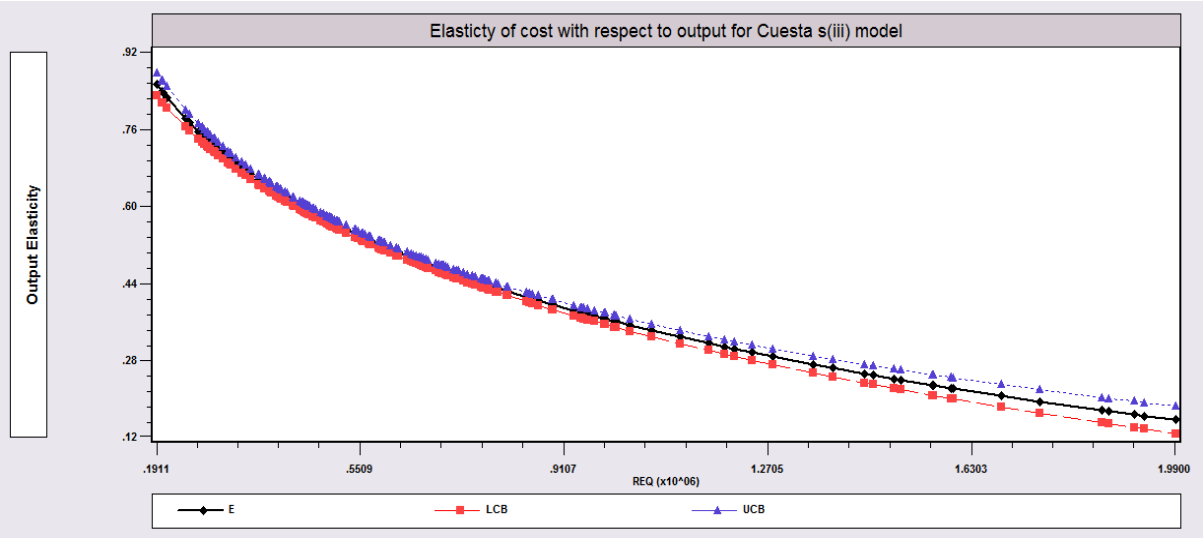

Fig. 3: Elasticity of cost with respect to output for Cuesta s(iii) model

Note to Figure 3: LCB – lower confidence bound, UCB – upper confidence bound. Requests are varied, all other variables are held at the sample mean.

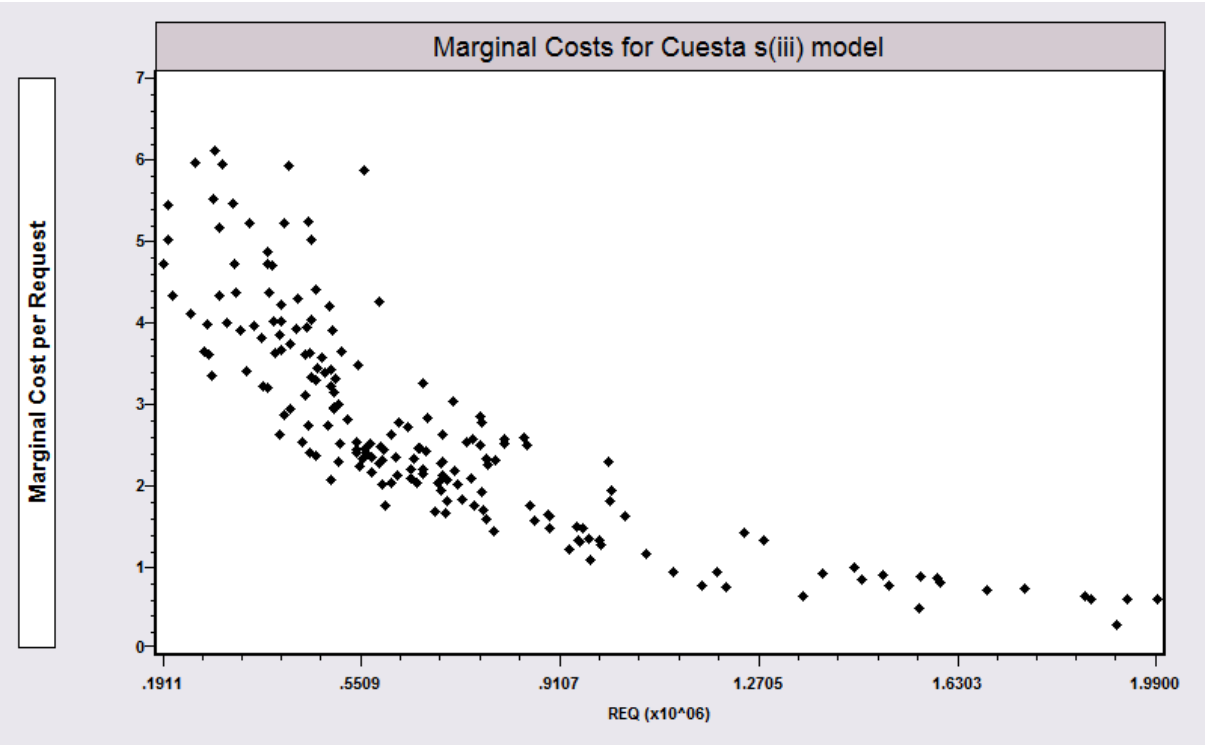

Fig. 4: Marginal cost (MC) for Cuesta s(iii) model

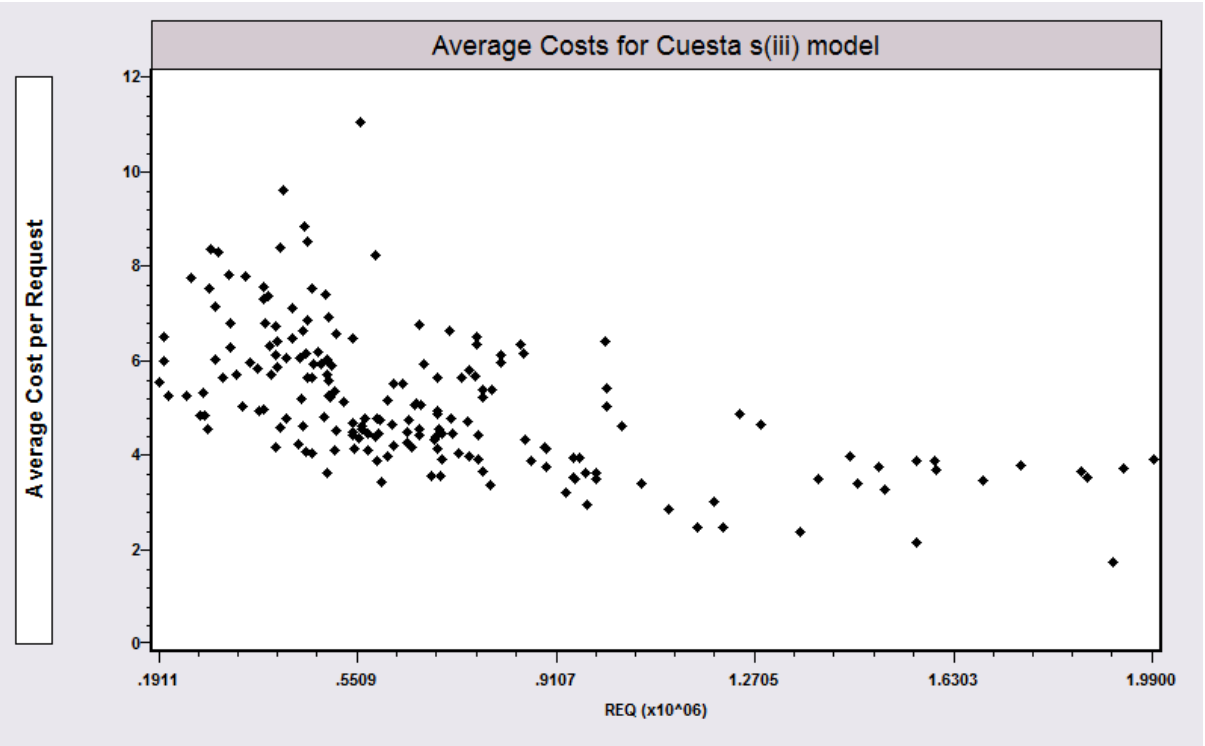

Fig. 5: Average cost (AC) for Cuesta s(iii) model

V. Discussion

Cost function parameters

This section draws on all models to examine the parameters of the cost function. The parameter estimates in the frontier models show reasonable concordance with each other and with the REM model, giving us confidence in our models.

The coefficient on input prices appears to be highly significant and in the range of 0.52-0.89, aside from two models, the Cuesta s(iii) and the TRE s(iii), which have values of 1.30 and 1.04, respectively. These estimates appear to be out of line with the remaining estimates. If the value of this coefficient was truly greater than 1, it would imply that operating costs were rising more quickly than input prices. However, we note that the 95% confidence intervals for both of these estimates include 1, meaning that we are unable to confirm that estimate of the coefficient, based on either of these models, exceeds 1. Of course, we only have data for labour input prices, and are thus unable to impose linear homogeneity of degree one on input prices, which gives rise to the possibility of beta estimates in excess of 1. We emphasise that the remaining models, including our benchmark REMs (which do not impose the distributional assumptions of the SF models), all appear to have estimates of the coefficient on labour input prices within a plausible range. Lastly, we note that other studies have shown large labour cost shares for biochemistry operating costs - approximately 80-90% (Department of Health, 2008 pp.44). This may explain the reported coefficients.

The time trend coefficients suggest a reduction in real laboratory operating costs of 0-2% per year. The 0-2% figure can then be seen as the shifting of the frontier over time. The frontier may exhibit downward shift if, for example, productivity in pathology production is increasing, which would support the findings of Holland et al., (2012).

Moving to the observable heterogeneity parameter coefficients (s(ii) variables), there was no clear finding of the impact of GP tests (the parameter was not statistically significant). From the healthcare commission (2007), a negative coefficient value was expected because primary care tests are thought to be cheaper than other tests.

The tests to requests ratio coefficients are in line with a priori expectations (positive and less than 1) from the literature (table 1). The estimated elasticity from this sample is in the range 0.12-0.48. The implications depend on the interpretation of this practice – it may be considered gaming by laboratories to inflate their performance figures; on the other hand it may be a reflection of a better quality of service since more patient information is being supplied per request.

The type of laboratory is found to be a source of cost heterogeneity, which matches previous literature (table 1). In our analysis, we were able to investigate this issue further. Laboratories situated in metropolitan areas are on average 9-17%[24] more costly than laboratories in rural areas. The findings for urban-based laboratories are that on average they are 0-5% more costly than rural laboratories. We caveat this finding by noting that the coefficient was significant in only three of fifteen models.

The foundation status of the host trust appears to be associated with a 4-10% reduction in operating costs for pathology laboratories. From the literature, profit incentives motivate hospitals to reduce costs to a greater extent than non-profit hospitals (Sloan, 2000), which is the aim of granting foundation status to a trust and should mean pathology services act commercially (Healthcare Commission, 2007).

Lastly, laboratories which provide teaching activities are found to have higher operating costs, in the range 0-5%, to those which do not; coefficients in only three of ten models were

[24] Because our model is estimated in logarithms, we have applied an exponential retransformation to recover our estimate of the effect on costs. To illustrate, for the Cuesta s(iii) model, exp(0.16) = 1.17, meaning that the beta on TYPE: Metropolitan from this model implies that costs are 17% higher than for TYPE: Rural laboratories.

statistically significant. This is in line with expectations, firstly because of the activity itself, but also because pathology services which are more specialised (and generally more expensive) tend to be associated with teaching activities, which may also be driving costs up (Department of Health, 2006). Moreover, this finding is in line with other health care studies (Gutacker et al., 2013).

The unobservable heterogeneity variable parameters (s(iii)) suggest that laboratories in London and the South are in the range 0-15% (statistically significant in 3 of 5 models) and 0-5% (statistically significant in 2 of 5 models), respectively, less expensive than laboratories from the North (the omitted dummy); and that operating costs of laboratories in the Midlands are on average 8-11% higher than those of laboratories in the North. From the literature, unobservable heterogeneity amongst these laboratories likely derives from information systems adoption, network activity and peer contact (Department of Health, 2006; Healthcare Commission, 2007; Eijkenaar, 2013)[25].

Efficiency Estimates

Our efficiency estimates are based on results from our preferred efficiency model: Cuesta s(iii).

To calculate our estimates of the potential savings we use laboratories' efficiency estimates in their final observed year. We calculate the potential cost of production if each laboratory adopted best practice (of that observed in the sample, denoted by each laboratory's efficiency estimate). Then, we subtract this estimate from the observed costs of laboratories to yield the potential available savings. We find potential savings of £32.8m in our sample (average cost efficiency in final year = 0.86).

[25] According to anecdotal evidence from pathologists, these features are more prevalent in London and the South and thus are likely driving this variation in costs.

We extrapolate to NHS pathology services (that is, all laboratories outside this sample and all other remaining pathology disciplines), giving an estimate of £390m per year of potential savings available to contribute to the Nicholson Challenge. This is around double the savings estimate that was proposed in the grey literature based on a much smaller sample – extrapolated comparably - of around £250m (Department of Health, 2008). Recalling that this data is for biochemistry services - the most mechanised of the five major pathology disciplines - we envisage that our estimates may well underestimate the true level of inefficiency, since mechanised pathology services are more homogenous than other disciplines (Kiechle and Main, 2002). We thus conclude that this is more likely a minimum efficiency saving than a maximum, which underlines the importance of pathology services for policy makers if expenditure reduction is high on their agenda.

However, driving out inefficiency may be more of a challenge amongst the more heterogeneous disciplines, such as hystocytology. First, not all laboratories conduct these services, meaning that there are fewer opportunities to compare practice and share knowledge. Second, that there is a paucity of available data in these disciplines means that measuring inefficiency may be more challenging (Buckell et al., 2013).

The average efficiency score over time is decreasing slightly. However, we find that individual etas imply that some laboratories are becoming more efficient over time, some are constant over time, and some are becoming less efficient over time (Fig.2); many of the laboratory-specific etas were found to be statistically significant. Information on the efficiency profiles of the individual laboratories is a powerful output of this type of top-down benchmarking as it indicates where further attention needs to be focused to drive out efficiency improvements. As noted earlier, the approach used to model efficiency change over time has been applied in economic regulation in other sectors. We do not identify individual laboratories for confidentiality reasons.

Given that we have reduced efficiency over time and technical change (falling costs) as per the time trend coefficient in our preferred model (i.e. frontier shift), it is informative to compute the Total Factor Productivity (TFP) Index (Coelli et al., 2005) to give an overall account of pathology performance. However, we do not observe costs which include capital, nor an output mix effect, meaning that it would be inappropriate to describe our measure as a TFP index. We therefore define a Multi-factor Productivity (MFP) Index as our measure of overall pathology performance.

Year	Average cost efficiency	Cost efficiency index	Frontier Shift	Overall MFP Index	change MFP
2007	0.868	1	1	1	0
2008	0.839	0.967	1.014	0.981	-1.9%
2009	0.857	0.987	1.029	1.016	3.5%
2010	0.847	0.976	1.044	1.020	0.3%
2011	0.858	0.989	1.059	1.048	2.8%

Table 6: Multi-Factor Productivity pathology laboratories

As can be seen in table 6, the overall MFP for pathology is increasing over time, from 1.000 in 2007 to 1.048 in 2011. The annual change is positive for three of the years and negative for one year. Overall, MFP increases by 4.8% over the period of study. Thus, the small reduction in the efficiencies of laboratories away from the frontier is more than offset by the gains in costs by the efficient firms (the frontier shift), yielding the overall MFP increase.

Economies of size in pathology

Due to our measure of costs not incorporating capital charges, we are, strictly speaking, unable to interpret changes in the relationship between output and costs as economies of scale. Accordingly, we refer to ‘economies of size’, and interpret this as the way in which operating costs change across the output range.

The cost elasticity estimates with respect to output indicate economies of size properties in pathology production (Fig. 3). Further, MC is falling faster than AC (Figs 4 and 5), meaning that the elasticity is falling (Fig 3), so the extent of economies of size is increasing as the scale of production increases; this will continue as long as MC falls faster than AC. This suggests that the growing formation of local pathology networks may help to lower costs for laboratories where production is pooled, which corresponds to pathology analysis elsewhere (Kiechle and Main, 2002). Encouragingly, this is being recognised by policy makers at the top level (Department of Health, 2011). It is of course possible that the economies of size be exhausted at some point, though we cannot conclude that based on our sample[26].

With regard to comparisons with previous studies, a direct comparison with the economies of scale finding in the Department of Health study (Department of Health, 2008) is difficult given that our measure does not incorporate capital costs. However, it is not clear that their measure did either, given that no empirical results on this issue are presented. Although capital cost information is collected (Department of Health, 2008, pp. 37) their only analysis (of unit costs) presented does not include these costs (Department of Health, 2008, pp. 44, 46, 48, 49). Therefore, on this issue, our study appears to be the first to present empirical evidence.

Merged laboratories

Using our preferred model, Cuesta specification (iii), and equation (3), we were able to simulate the effects of mergers between small laboratories. We find that, if the smaller laboratories in the sample merged, the sum of the implied predicted costs would be approximately 17% lower than those previously incurred by these laboratories separately. This suggests that there are potential considerable cost savings available via laboratories

[26] We note that the AC curve appears to be flattening towards the extreme of the sample (Figure 5). However, given that MC remains lower than AC at this point, this must be being driven by factors other than size which are associated with higher costs when size increases. However, further research with different data would be needed to draw any conclusion on the point at which size economies are exhausted.

pooling production. Indeed, this estimate suggests that these potential savings are greater than those available through efficiency improvements.

While we consider this to be a useful indicative valuation of the effect of potential pooling, we attach a number of caveats to this estimate; this exercise is ultimately a stylised scenario. Firstly, we note that there is no consideration to the additional costs incurred through merging (e.g. the costs to transport samples, the costs of service delays, etc.). Second, we do not take into account any effect on the quality of the service, the effect of specialisation or the interaction with other hospital services. Thirdly, we do not consider whether these laboratories are contiguous, which could potentially be limiting to mergers. On the other hand, this estimate is based on a small number of laboratories merging, in practice there is no limit to the number that can merge. In addition, we have assumed pairwise mergers; it is, of course, possible that multiple laboratories will merge. Thus, based on the last two caveats, the potential savings could be even larger than estimated here (as long as the subadditivity of costs continues).

VI. Conclusions

We have applied econometric efficiency estimation techniques to an under-researched area in health care literature: pathology. In doing so, we have developed performance measurement in this field beyond existing indicators benchmarking techniques. We have found, having controlled for cross-unit heterogeneity, 13% inefficiency in pathology services in the NHS in England. If this is indicative of NHS pathology as a whole, there could be £390m per year of available savings from pathology to contribute to the Nicholson Challenge of NHS efficiency savings. In addition, we found that the pooling of production looks to induce substantial gains in pathology cost savings. If smaller laboratories merged their production, they could save around 17% in their operating costs.

We have found that overall efficiency in pathology has decreased over time. The particular method that we have adopted also allows the time paths of efficiency for individual laboratories to be studied. We have also found frontier shift which decreases costs over time. Overall, MFP for the laboratories in our sample has increased by around 5% between 2007 and 2011.

We have estimated the magnitudes of various drivers of laboratory costs which were identified from previous pathology studies. Some of these drivers have not previously been quantified (e.g. the costs of teaching or the effect of the host trust having foundation status). We have paid particular attention to the elasticity of cost with respect to output. We have found economies of size, which is encouraging from a policy perspective because local networks are being formed in pathology services which increase the scale of production. We note, however, that our measure of costs does not include a component of capital, and thus are findings are limited to this extent. We also note that, although discussed in previous studies, no empirical evidence has been presented in previous literature on this issue (Department of Health, 2006; 2008). Therefore, on this issue, our study appears to be the first to present empirical evidence.

We believe these findings are important to policy makers because it provides them with the evidence needed to make informed decisions on the allocation of resources and on the management of pathology services. The method that we have adopted highlights performance variation both between decision making units (in our case, pathology laboratories) and over time. It has been applied by economic regulators outside health as a means of driving out efficiency improvements and we consider that it also has the potential to be applied much more widely in the health sector.

Acknowledgements

The authors would like to thank Dr Phill Wheat for providing some useful code and Professor Claire Hulme for some useful comments on several drafts of this paper. This work was funded under an Innovation in Quantitative Methods scholarship provided by the University of Leeds. We dedicate this work to the memory of the late Professor Rick Jones.

References

Aigner, D., Lovell, C. A. K. & Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, **6**, 21-37.

Alvarez, A., Amsler, C., Orea, L. & Schmidt, P. (2006) Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics. Journal of Productivity Analysis, **25**, 201-212.

Amado, C. & Santos, S. (2009) Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. Health Policy, **91**, 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 at http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/quarterly-monitoring-report-kingsfund-jun13.pdf. (accessed 12th January 2013)

Arocena, P., Saal, D. & Coelli, T. (2012) Vertical and Horizontal Scope Economies in the Regulated U.S. Electric Power Industry. The Journal of Industrial Economics, **60**, 434-467.

Baltagi, B. (2008) Econometric Analysis of Panel Data. John Wiley & Sons, Chichester.

Baltagi, B. and Li, Q. (1990) A Comparison of Variance Components Estimators Using Balanced Versus Unbalanced Data. Econometric Theory, 6, 283-285.

Battese, G. E. & Coelli, T. J. (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics, **38,** 387-399.

Battese, G. and Coelli, T. (1992) Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India. Journal of Productivity Analysis, **3**, 153-169.

Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T. & Peter C, S. (2005) Performance indicators: good, bad, and ugly. Journal of the Royal Statistical Society: Series A (Statistics in Society), **168**, 1-27.

Breusch, T. and Pagan, A. (1980) The LM Test and its Application to Model Specification in Econometrics. Review of Economic Studies, **47**, 239-254.

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

Christensen, L. R. & Greene, W. H. (1976) Economies of Scale in U.S. Electric Power Generation. Journal of Political Economy, **84**, 655-676.

Coelli, T., Rao, D., O'Donnel, C. & Battese, G. (2005) An Introduction to Efficiency and Productivity Analysis, Springer, New York.

Cooper, Z., Gibbons, S., Jones, S. & McGuire, A. (2012) Does Competition Improve Public Hospitals' Efficiency? Evidence from a Quasi-Experiment in the English National Health Service. Centre for Economic Performance Discussion Paper 1125. LSE.

Cornwell, C., Schmidt, P., and Sickles, R. (1990) Production Frontiers with Cross Sectional and Time Series Variation in Efficiency Levels. Journal of Econometrics, **46**, 185-200.

Cots, W., Chiarello, P., Salvador, X., Castells, X. & Quentin, W. (2011) DRG-based hospital payment: Intended and unintended consequences. In: Diagnosis Related Groups in Europe. Open University Press, Maidenhead.

Cuesta, R. (2000) A Production Model With Firm-Specific Temporal Variation in Technical Inefficiency: With Application to Spanish Dairy Farms. Journal of Productivity Analysis, **13**, 139-158.

Daidone, S. & Street, A. (2013) How much should be paid for specialised treatment? Social Science & Medicine, **84**, 110-118.

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 at http://www.connectingforhealth.nhs.uk/systemsandservices/pathology/projects/nlmc/carterreview2006.pdf (accessed 12th January 2013)

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 (accessed 12th January 2013)

Department of Health. (2009) Consolidation of pathology services. London: Department of Health. Available at http://www.evidence.nhs.uk/qipp (accessed 15th January 2013).

Department of Health. (2011) The Operating framework for the NHS in England 2012/13. London: Department of Health. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/152683/dh_131428.pdf.pdf (accessed 12th January 2013)

Dismuke, C. & Sena, V. (1999) Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. Health Care Management Science, **2**, 107-116.

Drummond, M., Sculpher, M., Torrance, G., O'Brien, B. & Stoddart, G. (2005) Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press, New York.

Duan, N. (1983) Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association, **78,** 605-610.

Eijkenaar, F. (2013) Key issues in the design of pay for performance programs. The European Journal of Health Economics, **14**, 117-131.

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J. & Scott, A. (2009) Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. BMJ, **339**.

Farrell, M. J. (1957) The Measurement of Productive Efficiency. Journal of the Royal Statistical Society. Series A, **120,** 253-290.

Farsi, M., Filippini, M. & Kuenzle, M. (2005) Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes, Applied Economics, **37**, 2127-2141.

Farsi, M. & Filippini, M. (2008) Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland, Health Economics, **17**, 335-350.

France, N. & Francis, G. 2005. Cross-laboratory benchmarking in pathology. Benchamrking, **12**, 523-538.

Giuffrida, A. & Gravelle, H. (2001) Measuring performance in primary care: econometric analysis and DEA, Applied Economics, **33**, 163-175.

Goddard, M. & Jacobs, R. (2009) Using Composite Indicators to Measure Performance in Health. In: Smith, P., Mossialos, E., Papanicolas, I. & Leatherman, S. (eds.) Performance Measurement for Health System Improvement. Cambridge University Press, Cambridge.

Greene, W. (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics, **126**, 269-303.

Greene, W. (2008) The Econometric Approach to Efficiency Analysis. in Fried, H., Lovell, K., and Schmidt, S. (eds.) The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press, Oxford.

Greene, W. (2010). A stochastic frontier model with correction for sample selection. Journal of Productivity Analysis, **34**, 15-24.

Greene, W. (2012a) LIMDEP Econometric Modelling Guide. Econometric Software Inc., New York.

Greene, W. (2012b) LIMDEP version 10. Econometric Software Inc., New York.

Greene, W. (2012c) Econometric Analysis. Harlow: Pearson Education Limited.

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.

Hausman, J. (1978) Specification Tests in Econometrics Econometrica, **46**, 1251-1271.

Healthcare Commission. (2007) Getting results: Pathology services in acute and specialist trusts. London: Commission for healthcare audit and inspection. Available at http://www.bipsolutions.com/docstore/pdf/16479.pdf (accessed 12th January 2013)

Health Select Committee Public Expenditure. (2010) Thirteenth Report of the Session 2010-2012. Available at www.publications.parliament.uk/pa/cm201012/cmselect/cmhealth/1499/149902.htm (accessed 12th January 2013)

Holland, D. & Trigg, G. (2010) Clinical Biochemistry/Chemical Pathology Core Report. National Pathology Benchmarking Review 2009/2010. Keele University.

Holland, D. & Trigg, G. (2011) Clinical Biochemistry/Chemical Pathology Core Report. National Pathology Benchmarking Review 2011/2012. Keele University.

Holland, D. & Trigg, G & Brookes, A. (2012) Clinical Biochemistry/Chemical Pathology Core Report. National Pathology Benchmarking Review 2010/2011. Keele University.

Hollingsworth, B. (2003) Non-Parametric and Parametric Applications Measuring Efficiency in Health Care, Health Care Management Science, **6**, 203-218.

Hollingsworth, B. (2008) The measurement of efficiency and productivity of health care delivery. Health Economics, **17**, 1107-1128.

Hollingsworth, B., Dawson, P. & Maniadakis, N. (1999) Efficiency Measurement of Health Care: A Review of Non-Parametric Methods and Applications. Health Care Management Science, **2**, 161-172.

Hollingsworth, B. & Peacocok, S. (2008) Efficiency Measurement in Health and Health Care, Routledge, Abingdon.

Jacobs, R., Street, A. & Smith, P. (2006) Measuring Efficiency in Health Care. Cambridge University Press, Cambridge.

Jondrow, J., Knox Lovell, C. A., Materov, I. S. & Schmidt, P. (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, **19**, 233-238.

Kiechle, F. & Main, R. (2002) The Hitchhiker's Guide to Improving Efficiency in the Clinical Laboratory. American Association for Clinical Chemistry, Washington.

Kumbhakar, S. C. & Hjalmarsson, L. (1995) Labour-Use Efficiency in Swedish Social Insurance Offices. Journal of Applied Econometrics, **10**, 33-47.

Kumbhakar, S. & Lovell, C. A. K. (2000) Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

Lai, H.-P. & Huang, C. (2010) Likelihood ratio tests for model selection of stochastic frontier models. Journal of Productivity Analysis, **34**, 3-13.

Laudicella, M., Olsen, K. & Street, A. (2010) Examining cost variation across hospital departments–a two-stage multi-level approach using patient-level data. Social Science & Medicine, **71**, 1872-1881.

Lester, H. & Roland, M. (2009) Performance Measurement in Primary Care. In: Smith, P., Mossialos, E., Papanicolas, I. & Leatherman, S. (eds.) Performance Measurement for Health System Improvement. Cambridge University Press, Cambridge.

Liebmann, R. (2011) Key performance indicators in pathology - why now? The Bulletin of The Royal College of Pathologists, **155**, 174-175.

Longo, R., Hulme, C. & Smith, A. (2012) Measures of Performance in Primary Care - a Review of Efficiency Studies. AUHE working paper series: WP 12_03. University of Leeds.

Manning, W. G. (1998) The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics, **17,** 283-295.

Marini, G., Miraldo, M., Jacobs, R. & Goddard, M. (2008) Giving greater financial independence to hospitals—does it make a difference? The case of English NHS Trusts. Health Economics, **17,** 751-775.

Meeusen, W. & Broeck, J. V. D. (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. International Economic Review, **18**, 435-444.

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

Mutter, R., Wong, H. S. & Goldfarb, M. (2008) The Effects of Hospital Competition on Inpatient Quality of Care. Inquiry, **45**, 263-279.

National Audit Office. (2012) Progress in making NHS efficiency savings. Available at http://www.nao.org.uk/publications/1213/nhs_efficiency_savings.aspx (accessed 12th January 2013)

NHS. 2013. The NHS belongs to the people: A Call to Action - The Technical Annex. London: NHS England.

NHS Confederation. (2010) Clinical responses to the downturn – pathology. Available at http://www.nhsconfed.org/Publications/Documents/Pathology.pdf (accessed 12th January 2013)

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.

Palangkaraya, A. & Yong, J. (2013) Effects of competition on hospital quality: an examination using hospital administrative data. The European Journal of Health Economics, **14**, 415-429.

Pitt, M. and L. Lee, (1981) The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry. Journal of Development Economics, **9**, 43-64.

Price, C. 2005. Benchmarking in pathology medicine: are we measuring the right outcomes?. Benchmarking, **12**, 449-466.

Propper, C. & Wilson, D. (2003) The Use and Usefulness of Performance Measures in the Public Sector. Oxford Review of Economic Policy, **19**, 250-267.

Propper, C., Burgess, S. & Gossage, D. (2008) Competition and quality: Evidence from the NHS internal market 1991-9. Economic Journal, **118**, 138-170.

Roberts, A., Marshall, L. & Charlesworth, A. (2012) A Decade of Austerity: The Funding Pressures Facing the NHS from 2011/2012 to 2021/22, London, The Nuffield Trust.

Available at http://www.nuffieldtrust.org.uk/sites/files/nuffield/121203_a_decade_of_austerity_full_report_1.pdf (accessed 12th January 2013)

Rosenman, R. & Friesner, D. (2004) Scope and scale inefficiencies in physician practices. Health Economics, **13**, 1091-1116.

Rosko, M. D. & Mutter, R. L. (2008) Stochastic Frontier Analysis of Hospital Inefficiency. Medical Care Research and Review, **65**, 131-166.

Siciliani, L., Sivey, P. & Street, A. (2013) Differences In Length Of Stay For Hip Replacement Between Public Hospitals, Specialised Treatment Centres And Private Providers: Selection or Efficiency? Health Economics, **22**, 234-242.

Sloan, F. (2000) Not-For-Profit Ownership and Hospital Behaviour. In Handbook of Health Economics. Elsevier, Amsterdam.

Smith, A. S. J. (2012) The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector. Transportation Research Part E: Logistics and Transportation Review, **48,** 503-515.

Smith, P. C. & 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**, 401-417.

Smith, P. and A. Street. (2006) Concepts and Challenges in Measuring the Performance of Health Care Organisations. The Elgar Companion to Health Economics, Edward Elgar, Cheltenham.

Smith, P. and A. Street. (2013) On The Uses of Routine Patient-Reported Health Outcome Data. Health Economics, **22**, 119-131.

Smith, S., Newhouse, J. P. & Freeland, M. S. (2009) Income, Insurance, And Technology: Why Does Health Spending Outpace Economic Growth? Health Affairs, **28**, 1276-1284.

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

Spinks, J. and Hollingsworth, B. (2009) Cross country comparisons of technical efficiency of health production: a demonstration of pitfalls. Applied Economics, **41**, 417-427.

Street, A. (2003) How much confidence should we place in efficiency estimates? Health Economics, **12**, 895-907.

Street, A., O'Reilly, J., Ward, P. & Mason, A. (2011) DRG-based hospital payment and efficiency: theory, evidence, and challenges. In: Diagnosis Related Groups in Europe. Open University Press, Maidenhead.

Szczepura, A., Davies, A., Fletcher, C. & Boussofiane, A. (1993) Efficiency and Effectiveness in General Practice. Journal of Management in Medicine, **7**, 36-47.

Valenstein, P., Praestgaard, A. & Lepoff, R. (2001) Six-year trends in productivity and utilization of 73 clinical laboratories: a College of American Pathologists Laboratory Management Index Program study. Arch Pathol Lab Med, **125**, 1153-61.

Veronesi, E., Mambretti, C. & Gazzaniga, P. (1997) Health care expenditure, laboratory services and IVD market. Int J Biol Markers, **12**, 87-95.

Vuong, Q. H. (1989) Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica, **57**, 307-333.

Wheat, P. & Smith, A. (2008) Assessing the Marginal Infrastructure Maintenance Wear and Tear Costs for Britain's Railway Network. Journal of Transport Economics and Policy, **42**, 189-224.

Wheat, P. & Smith, A. (2012) Is the choice of (t-T) in Battese and Coelli (1992) type stochastic frontier models innocuous? Observations and generalisations. Economics Letters, **116**, 291-2.