STOCHASTIC FRONTIER ANALYSIS OF SPECIALIST SURGEON CLINICS

STEVEN F. KOCH[†] AND JEAN D. SLABBERT[‡]

ABSTRACT. Stochastic Frontier Analysis was used to estimate the technical efficiency of specialist surgeon practices based in Gauteng Province, South Africa. The analysis was conducted for both single and multiple output production functions, while efficiency was allowed to depend upon surgeon and practice characteristics. Multiple output models, due to an increase in the number of observations, can be more more precisely estimated, and, since there are multiple observations per surgeon, can be estimated with fixed effects. The results of the analyses suggest that efficiency averages around 50% for this sample, and is convex in years of surgical experience. The benefit of multiple output analysis - improved precision - obtains, while surgeon level fixed effects alleviate some concerns related to unobserved heterogeneity.

Date: August 2011.

Key words and phrases. Stochastic Frontier Analysis, Technical Efficiency.

The authors would like to thank the South African Medical Association, the Foundation for Professional Development, Dr Gustav Wolvaardt, and the participating physicians for their help with this research. Any remaining errors are the sole responsibility of the authors.

1. INTRODUCTION

The South African Department of Health has, once again, thrown its support behind a national health insurance plan. Although the details of such a plan are only now being made available, one implication is that the private sector will be expected to provide care to a much larger number of patients than it does currently. However, it is not clear whether the private sector can accommodate more patients. One way to provide an estimate of the private sector's potential excess capacity is through the analysis of private sector health care delivery, which we undertake, below.

The private sector, at least the South African private sector, has, to the best of our knowledge, not been previously studied.¹ Therefore, this analysis contributes by providing one of the first examinations of efficiency in the private sector in South Africa. Specifically, we examine productive efficiency in a component of the South African private healthcare delivery sector, specialist surgeons, using data collected through the support of a number of health professional bodies. We estimate stochastic frontier production functions, from which we extract efficiency. As there are a number of different measures of output, there are a number of efficiency estimates that are comparable across measures of output. Given the small sample size that is available, we make a further contribution by exploring the potential benefits of a pooled multiple output analysis. We discuss whether or not pooling the data yields similar estimates of efficiency, while providing additional leverage with regard to the precision of the estimates and our ability to control for unobserved heterogeneity at the practice level.

We find that efficiency in our sample of specialist surgeons is not exceptionally high, averaging about 50% for the measures of output available in the data, suggesting that there is ample unused capacity in the delivery of, specifically, private surgical procedures. However, there is some variation across output measures. In

¹Previously published studies focus on the public sector. See for example, Kirigia, Lambo & Sambo (2000), Kirigia, Sambo & Scheel (2001), Zere, McIntyre & Addison (2001), as well as Kibambe & Koch (2007).

other words, some surgeons are better able to produce certain types of output than others. Furthermore, as we expect, experience is shown to improve efficiency, although at a decreasing rate. We also find that, in this sample, pooling the output measures yields comparable estimates of efficiency as in the single output models, and, therefore, precision can be improved through the estimation of a multiproduct production relation. However, there are some concerns related to the data that must be addressed. In particular, response data is voluntary, such that we might expect that relatively more (or less) organized or more (or less) efficient surgeons were able to complete the surveys, which would impact on our efficiency estimates and our conclusions. Therefore, future research needs to focus on the collection of more and better data, from which more precise estimates of production and productive efficiency can be gleaned.

The paper continues in Section 2 with a review of some of the literature related to efficiency analysis in the health care sector. The stochastic frontier analysis model is outlined in Section 3. Section 4 contains a description of the survey and data used in the analysis. The empirical results and some sensitivity analyses are discussed in Section 5, while Section 6 concludes and provides directions for future research.

2. BACKGROUND

Detailed reviews conducted by Hollingsworth (2003), Hollingsworth, Dawson & Maniadakis (1999) and Worthington (2004) highlight the large amount of research available on health sector efficiency, although it has focused primarily upon hospitals located in developed countries, and are either based on Data Envelopment Analysis (DEA) or Stochastic Frontier Analysis (SFA). Normally, SFA is used for single measures of costs or output, while DEA is used for multiple measures of

costs or output; below, however, we consider SFA within a multiple measure context. Although hospitals in developed countries receive the brunt of the attention, physicians, clinics and specialists have not been ignored.²

Further, there has been a recent uptick in research focusing on healthcare delivery in developing countries. Of particular relevance to our research is analyses conducted in Zambia, Kenya, Namibia and Sierra Leone - four African countries ranked similarly to South Africa by Evans, Tandon, Murray & Lauer (2000). For comparison, efficiency in Angola, China, Ghana and Turkey, developing countries ranking higher than South Africa, is also briefly considered.

Masiye, Kirigia, Emrouznejad, Sambo, Mounkaila, Chimfwembe & Okello (2006) use DEA to examine 40 health centres in Zambia that were approximately equally split between public and private ownership. Average technical efficiency was 83%, while allocative efficiency averaged 88%; however, 29% of private facilities were allocatively efficient, compared to just 4% of public centres. In another Zambian study, Masiye (2007) finds 67% average technical efficiency. In Kenya, Kirigia, Emrouznejad & Sambo (2002) find 26% of hospitals to be technically inefficient, but only about 16% inefficiency. Further, Kirigia, Emrouznejad, Sambo, Munguti & Liambila (2004), who also examine Kenya, find average efficiency levels around 65%. Although neither the Zambian nor the Kenyan studies are representative of their entire country's health sector, the estimated efficiency performance is generally better than it is in the South African analyses conducted by Kirigia et al. (2000), Kirigia et al. (2001), Zere et al. (2001) or Kibambe & Koch (2007).³

Additional analysis of healthcare delivery efficiency, primarily via DEA, has been conducted in Namibia and Sierra Leone. Zere, Mbeeli, Shangula, Mandlhate, Mutirua, Tjivambi & Kapenambili (2006) apply DEA to a Namibian sample; their

²See, for example, Chilingirian & Sherman (1990), Chilingirian (1993), Chilingirian (1995) for an analysis of doctors. Halsteinli, Kittlesen & Magnussen (2001) examine psychiatric clinics, while Schmacker & McKay (2008) provide evidence related to military hospital performance.

³Kirigia et al. (2001) find that inputs could be reduced by approximately 30%, and still keep output constant, while Zere et al. (2001) find that efficiency ranges from 68-74% in their sample of public hospitals. Kibambe & Koch (2007) find a much wider range of efficiency scores, due to the wide of assumptions applied, some of which are flexible enough to result in efficiency scores that are higher, on average, than the aforementioned studies in Kenya and Zambia.

technical efficiency estimates range between 62.7% and 74.3% from 1997 to 2001. Renner, Kirigia, Zere, Barry, Kirigia, Kamara & Muthuri (2005), who consider Sierra Leone district hospitals, arrive at an average technical efficiency score of 63%.

Research from other countries in Sub-Saharan Africa, which are ranked higher than South Africa by Evans et al. (2000), such as Angola and Ghana show mixed results. Estimates of technical efficiency in Angola are rather similar to the other African studies, while estimated efficiencies in Ghana are quite a bit higher. Kirigia, Emrouznejad, Cassoma, Zere & Barry (2008) consider Angola, finding technical efficiency ranging between 65.8% and 67.5% from 2000 to 2002. Ghana is examined by Osei, d'Almeida, George, Kirigia, Mensah & Kainyu (2005) and Akazili, Adjuik, Jehu-Appiah & Zere (2008), where average technical efficiency is much higher than in the other African countries considered. Osei et al.'s (2005) estimates average 81.4%, although 47% of the hospitals in their sample were technically inefficient, while Akazili et al.'s (2008) estimates average closer to 85%.

Surprisingly, the Chinese health sector, which is ranked more favourably than either Angola, Kenya, Namibia, South Africa or Zambia, appears to perform worse, at least on some of the comparisons that are available. Ng (2008) estimates that, overall, the Chinese health sector may be quite inefficient, averaging between 59.9% and 86.7% depending upon the year, location and output measure used in the analysis. Turkey, also ranked higher by Evans et al. (2000), appears to perform more poorly than some of the results noted for Zambia and Kenya. Sahin & Ozcan (2000), for example, find low levels of efficiency, averaging around 78.1%. In what follows, we provide further information on the efficiency of the South African healthcare sector, with special attention paid to the private sector, and compare our results to those already available in the literature.

3. METHODOLOGY

3.1. Surgical Production. The input data available for this analysis includes administrative staff, doctors and nurses, but is rather limited. There are only three multiple doctor practices, and each practice has either one nurse or zero nurses; therefore, the analysis is limited to single surgeon practices. However, those practices cover a range of specialties - orthopedic surgeons, vascular surgeons or other surgical specialties - which should also be included in the analysis. Furthermore, as there are only one or zero nurses, a nurse indicator variable is also included in the analysis. The lack of continuous input data, across all inputs, makes it impossible to consider a standard production function, such as the Cobb-Douglas. The basic production function considered in the analysis is, therefore, a modification of the standard Cobb-Douglas framework, as in (1), where i represents the specialist, $j = \{1, 2, 3\}$ represents the measure of output considered (total patients, new patients and surgeries), ℓ denotes the category associated with the dummy variables for specialty and nurse availability, D_ℓ , and a_i measures the number of administrative staff.⁴

$$(1) \quad \ln q_{ij} = \ln A_j + \sum_{\ell} \gamma_{\ell j} D_{\ell i} + \beta_j \ln a_i + \epsilon_{ij}$$

In addition to considering the preceding regressions, the analysis is extended in two different ways. Firstly, within each separate equation for output j , technical efficiency, discussed below, is allowed to depend on the length of service the surgeon has given in private practice. Secondly, the data is pooled - or stacked - over output measures, which allows for a number of potential benefits. For example, pooling artificially creates a greater number of observations, which increases the precision of the estimates. Furthermore, pooling creates multiple observations per surgeon, although only three, such that surgeon-level fixed effects can also be considered.

⁴One reviewer raised the concern that new patients ought to be a subset of total patients, and, therefore, the total patients output measure should be modified to measure net new patients. In analyses that is available upon request, we did consider the modification, finding only minor differences in the results; therefore, we retain the focus, as it is here, on total patients, new patients and surgeries.

The pooled model is described in (2), where J , a dummy variable, categorizes the measure of output.

$$(2) \quad \ln q_i = \ln A + \sum_{\ell} \gamma_{\ell} D_{i\ell} + \sum_j \psi_j J_{ij} \beta \ln a_i + \epsilon_i$$

3.2. Stochastic Frontier Analysis. The goal of Stochastic Frontier Analysis, as the moniker suggests, is to calculate an economically relevant frontier, such as a production frontier or a cost frontier, while accepting that economic data tends to be noisy, or stochastic. For example, consider the production function in (1). If such a function were estimated via OLS, it would yield a production function that was located in the middle of the data points, and, therefore, would not be a production frontier, where all of the production possibilities were contained by the frontier. Stochastic Frontier Analysis improves upon the OLS averaging process through the decomposition of the OLS error into a statistical noise component and a technical efficiency component.

The stochastic frontier model considered is underpinned by Battese & Coelli's (1995) extension of Aigner, Lovell & Schmidt's (1977) stochastic frontier model. In this error components model, the error term is decomposed into two parts, statistical noise and technical inefficiency, such that $\epsilon_{ij} = v_{ij} - u_{ij}$. The first component, statistical noise, is assumed to follow a normal distribution, $v_{ij} \sim \mathcal{N}(0, \sigma_{vj}^2)$. The second component, technical inefficiency, is assumed to follow a truncated normal distribution, $u_{ij} \sim |\mathcal{N}(\mu_j, \sigma_{uj}^2)|$. For identification purposes, u_{ij} and v_{ij} are assumed to be identically and independently distributed. The modeling emphasis is placed on $\mu_j = E[u_{ij}|\epsilon_{ij}]$. Jondrow, Materov, Lovell & Schmidt (1982) provide an estimator for that expectation, given in (3), based on the preceding assumptions.

$$(3) \quad E[u_{ij}|\epsilon_{ij}] = \frac{\sigma_j \lambda_j}{1 + \lambda_j^2} \left[\frac{\phi(a_{ij})}{1 - \Phi(a_{ij})} - a_{ij} \right]$$

However, the terms in (3) require additional elaboration.

$$(4) \quad a_{ij} = \frac{\mu_j}{\sigma_j \lambda_j} - \frac{\epsilon_{ij} \lambda_j}{\sigma_j}$$

In (3) and (4), $\sigma_j = \sqrt{\sigma_j^2} = \sqrt{\sigma_{uj}^2 + \sigma_{vj}^2}$, $\lambda_j = \sigma_{uj}/\sigma_{vj}$, $\phi(\cdot)$ is the normal density function and $\Phi(\cdot)$ is the normal distribution function.

In addition to decomposing the error, it is also possible to specify technical efficiency as a function of additional covariates. Battese & Coelli (1995) assume, as we do, that μ_j is a linear function of a number of explanatory variables. Specifically, in addition to (1) or (2), as well as (3), we assume that the underlying mean of the efficiency parameter can be further disaggregated.

$$(5) \quad \mu_j = E[u_{ij}|\epsilon_{ij}] = z_i\delta_j$$

In the single output analysis, z_i is a quadratic measure of the experience of the physician, based on the number of years in private practice. In the pooled analysis, surgeon-level fixed effects are also included in z_i .

The model is estimated, via maximum likelihood, using the “frontier” package, Coelli & Henningsen (2010), and it is estimated in R, R Development Core Team (2009). For ease of interpretation, production is measured in its natural log, see (1), such that inefficiency, contained in u_{ij} , measures the percentage deviation, actually the percentage reduction, between observed production and the production frontier.

4. DATA

A purposive survey collected in 2007, and reported in Slabbert (2010) forms the basis for the empirical analysis and results reported below. The focus of the survey was specialist surgeons, and, therefore, the survey was sent via mail to all Gauteng registered specialist surgeons.⁵ The survey, which was both voluntary and anonymous, queried surgeon and clinic characteristics, including experience, staff, patient amenities, patients, surgeries performed, and many other characteristics. This analysis focuses only upon single surgeon practices and makes use of information related to the surgeon’s inputs and outputs.⁶

⁵For a detailed discussion of the survey as well as problems surrounding data collection, see either Slabbert (2010) or Koch & Slabbert (2011).

⁶An analysis of practice income, costs and profits is undertaken elsewhere; see Koch & Slabbert (2011).

The South African Medical Association (SAMA), as well as the Foundation for Professional Development (FPD) sponsored the data collection efforts. The survey instrument supported by SAMA and FPD was broken into four major components. In the first component, the practice and patient profile were attended to, and, thus, the questions were directed towards practice size, attention to patient comfort, the number of patients, consultation length, number of surgeries performed, and many other items. In the second component, practice expenditures over the past month were requested. The third component addressed the personal and professional profile of the survey respondent, while the fourth component focused on the most sensitive information related to practice revenues.

Initially, 260 specialist physicians were requested to complete the confidential survey. However, only 69 did respond. Despite a response rate of 26.5%, which is less than we would like, it is better than the response rate achieved by Brentnall (2007). She only received responses from 5% of her sample, although she did draw from a much larger urn. The low response rate does raise concerns regarding the representativeness of the data, as well as the ability to generalize our results. Given that only eight respondents (11.5%) had been in private practice for five years or less, whereas 37 (53.6%) had been in private practice for fifteen years or more, it is likely that the respondents are not generally represented across the profession.⁷ Further, only two female surgeons responded, while all but five respondents were white.

Possibly the greatest concern that might arise, when considering efficiency using voluntary responses, is that efficiency could be correlated with the probability of response. For example, it might be true that only the most efficient manage to respond, which would lead to overestimates of efficiency. On the other hand, it might be true that only those with time on their hands, because they do less than they could, are likely to respond, which would result in understated estimates of

⁷However, it would also be reasonable to believe that specialist migration, “brain drain”, could have been highest amongst the youngest practitioners, such that the responses are representative of the South African specialist surgeon population.

efficiency. Unfortunately, we do not have data that allows us to instrument for the response probability, since we only have respondent data. Therefore, our reported results cannot be generalized to the population of specialist physicians practicing in Gauteng or beyond, and should be understood in that context.

The analysis of efficiency focused on information gleaned from the first and third survey components. In particular, we limited our data to single surgeon practices, but also made use of information on the number of nurses, the number of administrative staff, the total number of patients, the number of new patients and the number of surgeries performed. As previously noted, there were 69 respondents; however, due to missing data for various inputs or outputs, the analysis sample was trimmed to 58. Descriptive statistics of the data used in the analysis are reported in Table 1.

TABLE 1. Descriptive Statistics

Output Measures			Input Measures		
	Mean	Std. Dev.		Mean	Std. Dev.
Total Patients	203.56	120.2	Nurse (=1 if positive)	0.33	0.5
New Patients	70.82	39.9	Administrators	2.49	1.3
Surgeries	37.84	26.5	Orthopedists (=1)	0.64	0.5
			Vascular Surgeons (=1)	0.21	0.4
			Years in Practice	18.39	10.3

5. EMPIRICAL RESULTS

In this section, we present the primary results related to the stochastic frontier estimates, and error decompositions described in equations (1) or (2), (3) and (5). Although the main focus of the analysis is on the estimation of efficiency for each of the responding surgeons in our survey, production function estimates are also reported and interpreted. Three sets of analyses are considered. In the first, each input is estimated separately, while efficiency is assumed to be based on a truncated normal distribution. In the second, each input is, again, estimated separately; however, efficiency is allowed to depend upon additional factors, such that the truncated normal distribution assumption is no longer maintained. In the third, the

inputs are stacked, allowing for a seemingly unrelated regression of the production function. In this last analysis, efficiency is assumed to follow a truncated normal, and it is also allowed to depend on other factors. Within each of these analyses, additional considerations related to the similarity of efficiencies across outputs, as well as the similarity of efficiency rankings across outputs, are also discussed.

5.1. Single Output Production Functions.

5.1.1. *No Efficiency Controls.* The results from the initial stochastic frontier analysis are presented in Table 2. In terms of production, output, regardless of the measure used, is primarily determined by the coefficient of technology, which is, in turn, affected by the practice's surgical specialty and whether or not the practice employs any nurses. The total number of patients is approximately 35% higher, if there is a nurse available in the practice, while practices employing a nurse deal with approximately 53% more new patients than those without a nurse. For surgeries, the corresponding estimate is 69%. In terms of surgical specialties, only total patients are practice-specific. Vascular surgeon patient rolls are nearly 49% larger than for any other surgeon, while orthopedic surgeons practices serve 15% more patients than any other type of surgeon. The input elasticity for administrative staff, like the specialty estimates, is only significant for total patients, and is 0.58. Unsurprisingly, larger practices require a larger support staff.

Economically, although nurses are treated categorically, since there are only non-nurse and single nurse practices, a rough estimate of the returns to scale for the practice production functions can be calculated as the sum of the administrative input elasticity and the nurse input semi-elasticity. For total patients, that sum is 0.93; for new patients, the sum is 0.63, and for surgeries, 0.93. Assuming normality and independence across the inputs, which is a reasonable simplification given the linear projection properties of the regression, the input elasticity standard errors are approximated to be (0.01), (0.32) and (0.31) for total patients, new patients and surgeries, respectively. In other words, total surgeries are produced with decreasing returns to scale, although those returns are nearly constant, while decreasing, zero

and constant returns to scale can not be rejected for either new patient or surgery production.

TABLE 2. Single Input Efficiency Estimates and Truncated Normality

Variable	Total Patients	New Patients	Surgeries
Intercept	4.902 ^a (0.01)	4.439 ^a (0.55)	3.392 ^a (0.34)
Administration	0.582 ^a (0.01)	0.105 (0.29)	0.236 (0.24)
Nurse	0.300 ^a (0.01)	0.428 ^a (0.12)	0.527 ^b (0.20)
Orthopedic	0.153 ^a (0.01)	0.321 (0.27)	0.303 (0.22)
Vascular	0.402 ^a (0.01)	0.206 (0.31)	0.322 (0.30)
Error Decomposition			
σ^2	1.482 ^a (0.39)	0.727 ^a (0.13)	5.649 ^c (2.34)
λ	1.000 ^a (0.00)	1.000 ^a (0.00)	0.978 ^a (0.02)
μ	-5.612 (0.57)	0.497 (0.45)	-4.700 ^d (2.61)
Performance Measures			
Average Efficiency	0.54	0.49	0.55
$\ln \mathcal{L}$	-43.3	-46.7	-67.6

Source: Simultaneous estimates of (1) and (3); outputs estimated separately.

Standard errors in parentheses. ^a Significant at 0.001. ^b Significant at 0.01.

^c Significant at 0.05. ^d Significant at 0.1. $n = 58$.

Efficiency is estimated to be very low in the sample, nearly 50% for these surgical practices, regardless of which measure of output is used. These efficiency estimates are lower than nearly all of the estimates from Africa and South Africa that are available in the literature. The estimates are lower than South African estimates in Zere et al. (2001) and Kirigia et al. (2001), although they present their estimates in a different fashion. If comparisons were drawn with other research conducted in Africa, our estimates are also lower than estimates from Zambia, Masiye (2007), Kenya, Kirigia et al. (2004), Namibia, Zere et al. (2006), and Sierra Leone, Renner et al. (2005). In other words, our estimates suggest that private surgical practices in South Africa are less efficient than public clinics and hospitals in Africa.

In each of the analyses, the estimated mean from the truncated normal distribution, recall the assumption $u_{ij} \sim |\mathcal{N}(\mu_j, \sigma_{uj}^2)|$, ranges from a significant -5.61 to an insignificant 0.50, suggesting that production is generally lower for the inefficient firms, as it should be. Furthermore, the ratio of signal to noise, as defined by λ_j , is approximately 1.00, regardless of the measure of output, and is statistically significantly different from zero. Finally, the estimate of the variance of the composite error term suggests that this variance is statistically different from zero, implying that the distributions of noise and inefficiency are reasonably well defined in the analysis, i.e., the distributions are not degenerate.

5.1.2. Including Inefficiency Determinants. In the preceding analysis, a truncated normal distribution was assumed. In what follows that assumption is relaxed, in the sense that the underlying mean of the decomposed error term is allowed to be a function of additional determinants. As experience is likely to be an important determinant of productivity, a decision was made to include a quadratic measure of experience in a bid to explain inefficiency. The results of this analysis, an extension of the results contained in Table 2, are presented in Table 3.

Unlike the analysis in Kibambe & Koch (2007), the inclusion of additional controls does not have much impact on estimated average efficiency by product.⁸ However, the inclusion of those controls, which are significant for total patient inefficiency and new patient inefficiency, does impact the point estimates of returns to scale. For total patients, the returns to scale are 0.66 (s.e.≈0.18), while for new patients and surgeries, the respective estimates are 0.70 (s.e.≈0.37) and 0.71 (s.e.≈0.30), suggesting, in all cases, that decreasing, constant and increasing returns to scale remain plausible.

As alluded to above, mean efficiency estimates are expected to be negative. In a similar vein, the signs of the efficiency determinants are associated with inefficiency. Therefore, inefficiency, for both total and new patients is decreasing in years of

⁸Primarily, this is due to the fact that our analysis only allows inefficiency to be explained by additional flexibility, whereas, Kibambe & Koch (2007) allow for additional flexibility at the production level.

TABLE 3. Single Input Efficiency Estimates with Inefficiency Determinants

Variable	Total Patients	New Patients	Surgeries
Intercept	5.161 ^a (0.25)	4.347 ^a (0.42)	3.436 ^a (0.36)
Administration	0.459 ^a (0.14)	0.244 (0.28)	0.186 (0.24)
Nurse	0.196 (0.12)	0.463 ^c (0.24)	0.525 ^b (0.18)
Orthopedic	0.099 (0.15)	0.229 (0.30)	0.240 (0.24)
Vascular	0.076 (0.27)	0.057 (0.30)	0.200 (0.28)
Explaining Efficiency			
Intercept	1.858 ^b (0.63)	1.986 ^b (0.70)	-22.480 (69.12)
Years Private	-0.248 ^d (0.13)	-0.184 ^b (0.07)	-0.191 (1.11)
Yrs Private Sq.	0.006 ^d (0.00)	0.004 ^b (0.00)	0.018 (0.04)
Error Decomposition			
σ^2	0.830 ^d (0.46)	0.565 ^c (0.27)	12.823 (34.54)
λ	0.982 ^a (0.03)	1.000 ^a (0.00)	0.989 ^a (0.03)
Performance Measures			
Average Efficiency	0.55	0.49	0.59
$\ln \mathcal{L}$	-41.6	-40.0	-62.3

Source: Simultaneous estimates of (1), (3) and (5); outputs estimated separately.

Standard errors in parentheses. ^a Significant at 0.001. ^b Significant at 0.01.

^c Significant at 0.05. ^d Significant at 0.1. $n = 58$.

experience, but increasing in its square, i.e., inefficiency is convex in experience. As with the previous analyses, the signal to noise ratio represented by λ is near unity. However, only the estimates for total patients and new patients can be properly referred to as estimates of stochastic frontiers, because the estimated variance for surgeries is degenerate.

5.1.3. Efficiency Rankings and Correlations. Given the fact that a number of different single input stochastic frontier models were considered, efficiency correlations and the correlations of efficiency rankings across each of the single input analyses were analyzed to provide an indication related to the potential to pool the models.

Along these lines, we first considered the correlation between estimated efficiency scores across each of the different output measures for both of the preceding single input analyses. The results are located in Table 4.

TABLE 4. Single Input Efficiency Correlations across Output Measures

	Without Decomposition		With Decomposition	
	Total Patients	New Patients	Total Patients	New Patients
New Patients	0.70		0.72	
Surgeries	0.46	0.42	0.49	0.42

Source: Authors' calculations.

Although the estimated correlations are reasonably high, they are not indicative of strong similarities across the various measures of output. Therefore, correlations of efficiency ranks were also examined. The ranking correlations are presented in Table 5.

TABLE 5. Correlations of Efficiency Ranking across Output Measures

	Without Decomposition		With Decomposition	
	Total Patients	New Patients	Total Patients	New Patients
New Patients	-0.16		-0.17	
Surgeries	0.09	-0.11	0.31	-0.06

Source: Authors' calculations.

The results in Tables 4 and 5 suggest that, although there is a reasonably strong correlation between efficiency scores across the various measures of output, the actual efficiency rankings are not very similar. In other words, the level of efficiency across outputs is reasonably consistent, even though the efficiency rank order of these surgeons is not consistent. Although these correlations do not, by themselves, guarantee that pooling the data is an appropriate endeavour, they provide us with a baseline correlation matrix that can be used to determine if pooling the data significantly alters the efficiency estimates.

5.2. Multiple Output Production Functions. Due to the limited number of respondents, the single output analysis tends to be associated with imprecise estimates. For that reason, multiple output production functions were estimated.

Since the right hand side variables are identical across outputs, the analysis can best be described as a pooled analysis, where the estimates are restricted to be the same across all input measures.

Pooling the data also provides an additional potential benefit, although, as will be discussed below, that benefit cannot be completely realized, due to the limited number of observations per surgeon. Specifically, although an attempt has been made to deal with practice heterogeneity through the inclusion of surgeon experience, other forms of heterogeneity are also likely to exist in the data. For example, the quality of the surgeon's training, the reputation of the surgeon and the profile of the patients are all unobservable. If patients visiting one physician are generally worse off than others, that physician will not be able to see as many. This *House* effect, referring to the TV series, would imply high levels of inefficiency for very good doctors, potentially overstating the degree of inefficiency in the sample.⁹ Pooling the data, therefore, allows us to control for surgeon level heterogeneity, via fixed-effects estimates, but does require the assumption that heterogeneity is constant across outputs.

5.2.1. Pooled Analysis. Four separate pooled analyses were estimated. In the first, referred to as Pooled 1 in Table 6, no attempt is made to control for either efficiency or surgeon level fixed effects. Pooled 2 controls for physician experience, but not fixed effects. In the final two columns, surgeon fixed effects were included to control for the determinants of inefficiency; however, in Pooled 3, experience was not included, while experience was included in Pooled 4. In order to control for potential differences in output measure, the type of output was included as a dummy variable to further explain the level of production. In all cases, these were significant and negative, as expected, since all surgeons perform fewer surgeries and have fewer new patients than they have total patients.

The first of the purported benefits of pooling are evident across each of the pooled analyses, as nearly all of the included controls are estimated precisely enough to be

⁹We would like to thank an anonymous reviewer for reminding us of the importance of case-mix heterogeneity.

TABLE 6. Multiple Input Efficiency Estimates and Error Decomposition

	Without Fixed Effects		With Fixed Effects	
Variable	Pooled 1	Pooled 2	Pooled 3	Pooled 4
Intercept	5.077 ^a (0.19)	5.158 ^a (0.18)	4.932 ^a (0.19)	4.941 ^a (0.18)
Administration	0.412 ^a (0.10)	0.382 ^a (0.10)	0.373 ^b (0.13)	0.386 ^b (0.13)
Nurse	0.277 ^a (0.08)	0.276 ^a (0.08)	0.299 ^b (0.11)	0.193 ^b (0.10)
Orthopedic	0.210 ^c (0.11)	0.192 ^d (0.11)	0.300 ^b (0.10)	0.281 ^c (0.12)
Vascular	0.219 ^d (0.12)	0.156 (0.13)	0.222 ^d (0.13)	0.207 (0.14)
NewPat	-1.027 ^a (0.09)	-1.018 ^a (0.64)	-1.023 ^a (0.07)	-1.021 ^a (0.07)
Surg	-1.560 ^a (0.11)	-1.556 ^a (0.10)	-1.686 ^a (0.08)	-1.676 ^a (0.09)
Explaining Efficiency				
Intercept	-2.485 (3.82)	1.287 ^c (0.64)	-168.06 (292.91)	115.490 (487.4)
Years Private		-0.263 ^c (0.10)		-22.829 (96.54)
Yrs Private Sq.		0.007 ^b (0.00)		0.528 (2.23)
Error Decomposition				
σ^2	3.332 (2.92)	1.500 ^c (0.60)	0.349 ^a (0.07)	0.349 ^a (0.06)
λ	0.988 ^a (0.01)	0.977 ^a (0.01)	0.834 ^a (0.05)	0.841 ^a (0.05)
Performance Measures				
Average Efficiency	0.53	0.52	0.62	0.61
$\ln \mathcal{L}$	-171.4	-159.7	-83.6	-83.5

Source: Simultaneous estimates of (2), (3) and (5); pooled outputs.

Standard errors in parentheses. ^a Significant at 0.001. ^b Significant at 0.01.

^c Significant at 0.05. ^d Significant at 0.1. $n = 184$.

statistically significant.¹⁰ Furthermore, the resulting estimates for returns to scale are more precisely estimated. Returns to scale are estimated to be 0.69 (s.e.≈0.12), 0.66 (s.e.≈0.12), 0.67 (s.e.≈0.17), and 0.58 (s.e.≈0.16) for Pooled 1, 2, 3 and 4, respectively; three of the four are statistically and significantly less than unity, using

¹⁰Importantly, a casual comparison of parameters across Tables 2, 3 and 6 points to extremely similar estimation results.

conventional confidence intervals, suggesting that these surgical practices operate at decreasing returns to scale.

Unfortunately, the second of the pooling benefits is less apparent. Although a likelihood ratio test between either Pooled 1 or Pooled 2 and either Pooled 3 or Pooled 4 suggests that the fixed effects are jointly significant, there is rather obvious multicollinearity between the fixed effects and the efficiency intercept.¹¹ Furthermore, when practice level experience is also included with the fixed effects, see Pooled 4, similar multicollinearity problems can also be observed, while the effect of experience on inefficiency cannot be pinned down.

Despite the multicollinearity problems that do arise in the fixed effects analysis, the pooled analysis production function results are rather similar to each other, and not too dissimilar from the single input analyses. Therefore, one tentative conclusion that can be drawn from this analysis is that unobserved heterogeneity, at least of the type that is fixed across surgeons, is not a serious problem, at least in this sample of specialist surgeon's from Gauteng. However, it should be noted that this conclusion is based on the assumption that statistical noise and technical efficiency are independent of each other, for identification purposes.¹²

5.2.2. Efficiency Correlations. As with the single input analysis, and as a check of similarity across the single input and multiple input analysis, the pooled outcome efficiency measures were examined further. As in Table 4, Table 7 presents the correlations between the estimated efficiency from the pooled analysis. In this case, the efficiency estimates are taken from Pooled 1 and Pooled 2. Given the similarities in the stochastic frontier analysis across the various pooled models,

¹¹Multicollinearity is easily seen via the efficiency intercepts, which jump by a factor of close to 100, while the fixed effects, not shown, tend to be close to the estimated efficiency intercept or its negative. However, these fixed effects are jointly significant. In terms of the likelihood ratio test, comparing Pooled 2 with Pooled 3, for example, yields $2(-83.6 + 159.7) = 152.2$; the critical χ^2_{58} value is approximately 77 at 5% significance.

¹²Including fixed effects in the production function, rather than in the efficiency estimates does yield different estimates. However, the previously noted multicollinearity is more problematic at the production function level than at the efficiency explanation level. Specifically, we were unable to estimate the stochastic frontier model in all cases.

it is not surprising that the resulting correlations between efficiency are nearly identical.

TABLE 7. Multiple Input Efficiency Correlations across Output Measures

	Without Decomposition		With Decomposition	
	Total Patients	New Patients	Total Patients	New Patients
New Patients	0.79		0.79	
Surgeries	0.45	0.42	0.45	0.43

Source: Authors' calculations.

In comparing Tables 4 and 7, the correlations, although not identical, are quite similar, which provides additional support for pooling the data. However, further support is offered by considering the correlations between the multiple output and single output models. Table 8 contains the correlations between the single output efficiencies and Pooled 2 efficiencies.¹³ The important result to be observed is along the diagonal, which shows the degree to which the single output and multiple output efficiencies are correlated for the same output measure. Those correlations range from 0.94 to 0.97, suggesting that the multiple output analysis is broadly in line with the single output analysis. In other words, efficiency in the separate analysis corresponds well with estimated efficiency from the pooled analysis. Furthermore, cross-analysis correlations are very similar to the within-analysis correlations. In summary, pooling the data does not obviously affect the underlying efficiency scores. Therefore, pooling the data, which increases the total number of observations and, thus, improves precision in the underlying stochastic frontier regression, is a reasonable option in this limited sample.

TABLE 8. Efficiency Correlations Between Single and Multiple Output Analysis

Separate	Pooled Analysis		
Analysis	Total Patients	New Patients	Surgeries
Total Patients	0.97		
New Patients	0.76	0.97	
Surgeries	0.50	0.48	0.94

Source: Authors' calculations.

¹³Similar results are available for comparison across the other pooled models.

6. DISCUSSION AND CONCLUSION

In the preceding analysis, a variety of stochastic frontier analysis models has been considered in order to examine the technical efficiency of specialist surgeons practicing in Gauteng Province, South Africa. The responding surgeons are at operating at approximately 50% efficiency, although that estimate nears 60% in the pooled analysis. Implicitly, these specialists are able to take on more patients and more surgeries than they are currently, while using the same number of resources.

The results are broadly similar to previous analyses of, especially African, healthcare sector efficiency, in the sense that they suggest very low levels of efficiency in the healthcare sector. Research conducted in South Africa by, amongst others, Kirigia et al. (2000), Kirigia et al. (2001), Zere et al. (2001) and Kibambe & Koch (2007) have found that healthcare production in the South African public sector is rather inefficient. The results presented above, suggest that the healthcare production in the South African private sector is also inefficient, and that surgical production, at least in the private sector, is generally less efficient than general healthcare production in the public sector.

Although our results point to inefficiency in private healthcare delivery, it should be noted that there are a number of shortcomings in the preceding analysis. One major concern is the low response rate in our survey. Conservatively, it would be reasonable to assume that participation was more likely for those who were better organized (or more efficient), in which case, our efficiency estimates would be overstated. However, it would not be unreasonable to assume that participation was more likely amongst those not using their time to the fullest, in which case respondents might be more inefficient than our specialist surgeon sample.

Another concern, and one that arises in all examinations of health care production, is that our measure of output does not truly capture healthcare production, while unobserved heterogeneity is also potentially problematic. Although it would be preferable to have a measure of health improvement, rather than a simple stock of patients or flow of patients and surgeries, data to that effect is not available, and,

therefore, we only report on efficiency based on stocks and flows of patients and surgeries. Furthermore, although unobserved heterogeneity is potentially a problem, multiple output stochastic frontier analysis, based on pooled data and surgeon fixed effects, suggests that unobserved heterogeneity is not a serious problem in this sample.

From a policy perspective, our results lend some support to the notion that the private sector is capable of serving additional customers, such that access and use could be improved. However, given the shortcomings in our data, we do not think it is appropriate to conclude that specialty surgeons in Gauteng are wasting resources, even though our results suggest that resources could be more appropriately allocated. Future research in this area is warranted, and additional data is needed in order to conduct additional research. Further analysis is needed in order to create a firmer picture of the delivery of healthcare within the private sector, providing much better information to those interested in further developing healthcare delivery policy.

REFERENCES

- Aigner, D. J., Lovell, C. A. K. & Schmidt, P. (1977), 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics* **6**(1), 21–37.
- Akazili, J., Adjuik, M., Jehu-Appiah, C. & Zere, E. (2008), 'Using data envelopment analysis to measure the extent of technical efficiency in public centres in Ghana', *BMC International Health and Human Rights* **8**(11).
- Battese, G. E. & Coelli, T. J. (1995), 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics* **20**, 325–332.
- Brentnall, V. F. (2007), 'Exclusive survey - expenses: Rising costs hit all physicians', *Medical Economics* **84**, 26–31.
- Chilingerian, J. A. (1993), Exploring why some physicians' hospital practices are more efficient: Taking DEA inside the hospital, in A. Charnes, W. W. Cooper, A. Y. Levin & L. M. Seiford, eds, 'Data Envelopment Analysis: Theory, Methodology and Applications', Kluwer, Boston.
- Chilingerian, J. A. (1995), 'Evaluating physician efficiency in hospitals: A multivariate analysis of best practices', *European Journal of Operational Research* **80**, 548–574.

- Chilingerian, J. A. & Sherman, H. D. (1990), ‘Managing physician efficiency and effectiveness in providing hospital services’, *Health Services Management Research* **3**, 3–15.
- Coelli, T. & Henningsen, A. (2010), *frontier: Stochastic Frontier Analysis*, R package version 0.996-6 edn.
- URL:** <http://CRAN.R-project.org/package=frontier>
- Evans, D. B., Tandon, A., Murray, C. J. L. & Lauer, J. A. (2000), The comparative efficiency of national health systems in producing health, GPE Discussion Paper Series 29, World Health Organization.
- Halsteinli, V., Kittlesen, S. A. C. & Magnussen, J. (2001), ‘Scale, efficiency and organization in Norwegian psychiatric outpatient clinics for children’, *The Journal of Mental Health Policy and Economics* **4**, 79–90.
- Hollingsworth, B. (2003), ‘Non-parametric and parametric applications measuring efficiency in health care’, *Health Care Management Science* **6**, 203–218.
- Hollingsworth, B., Dawson, P. J. & Maniadakis, N. (1999), ‘Efficiency measurement of health care: A review of non-parametric methods and applications’, *Health Care Management Science* **2**, 161–172.
- Jondrow, J., Materov, I., Lovell, K. & Schmidt, P. (1982), ‘On the estimation of technical inefficiency in the stochastic frontier production function model’, *Journal of Econometrics* **2/3**, 233–238.
- Kibambe, J. N. & Koch, S. F. (2007), ‘DEA applied to a Gauteng sample of public hospitals’, *South African Journal of Economics* **75**(2), 351–368.
- Kirigia, J. M., Emrouznejad, A., Cassoma, B., Zere, E. & Barry, S. (2008), ‘A performance assessment method for hospitals: The case of municipal hospitals in Angola’, *Journal of Medical Systems* **32**, 508–519.
- Kirigia, J. M., Emrouznejad, A. & Sambo, L. G. (2002), ‘Measurement of technical efficiency of public hospitals in Kenya: Using Data Envelopment Analysis’, *Journal of Medical Systems* **26**(1), 39–45.
- Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Munguti, N. & Liambila, W. (2004), ‘Using data envelopment analysis to measure the technical efficiency of public health centers in Kenya’, *Journal of Medical Systems* **28**, 155–166.
- Kirigia, J. M., Lambo, E. & Sambo, L. G. (2000), ‘Are public hospitals in Kwazulu-Natal province of South Africa technically efficient?’, *African Journal of Health Sciences* **7**(3-4), 5–11.
- Kirigia, J. M., Sambo, L. G. & Scheel, H. (2001), ‘Technical efficiency of public clinics in Kwazulu-Natal province of South Africa’, *East African Medical Journal* **78**(3 Supp), S1–S13.

- Koch, S. F. & Slabbert, J. D. (2011), 'An analysis of specialist surgeons and their practices', *South African Journal of Economic and Management Sciences* **14**(3), forthcoming.
- Masiye, F. (2007), 'Investigating health system performance: An application of Data Envelopment Analysis to Zambian hospitals', *BMC Health Services Research* **7**(58).
- Masiye, F., Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Mounkaila, A., Chimfwembe, D. & Okello, D. (2006), 'Efficient management of health centres human resources in Zambia', *Journal of Medical Systems* **30**, 473–481.
- Ng, Y. (2008), 'The productive efficiency of the health sector of china', *The Review of Regional Studies* **38**(3), 381–393.
- Osei, D., d'Almeida, S., George, M. O., Kirigia, J. M., Mensah, A. O. & Kainyu, L. H. (2005), 'Technical efficiency of public hospitals and health centres in Ghana: A pilot study', *Cost Effectiveness and Resource Allocation* **3**(9).
- R Development Core Team (2009), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>
- Renner, A., Kirigia, J. M., Zere, E. A., Barry, S. P., Kirigia, D. G., Kamara, C. & Muthuri, L. H. K. (2005), 'Technical efficiency of peripheral health units in Pujehun district of Sierra Leone: A DEA application', *BMC Health Services Research* **5**(77).
- Sahin, I. & Ozcan, Y. A. (2000), 'Public sector hospital efficiency for provincial markets in Turkey', *Journal of Medical Systems* **24**, 307–320.
- Schmacker, E. R. & McKay, N. L. (2008), 'Factors affecting productive efficiency in primary care clinics', *Health Services Management Research* **21**, 60–70.
- Slabbert, J. D. (2010), Measuring efficiency in specialist doctor practices in Gauteng, South Africa: Stochastic frontier analysis, MCOM: Econometrics, University of Pretoria.
- Worthington, A. C. (2004), 'Frontier efficiency measurement in health care: A review of empirical techniques and selected applications', *Medical Care Research and Review* **61**(2), 135–170.
- Zere, E., Mbeeli, T., Shangula, K., Mandlhate, C., Mutirua, K., Tjivambi, B. & Kapenambili, W. (2006), 'Technical efficiency of district hospitals: Evidence from Namibia using Data Envelopment Analysis', *Cost Effectiveness and Resource Allocation* **4**(5).
- Zere, E., McIntyre, D. & Addison, T. (2001), 'Technical efficiency and productivity of public sector hospitals in the South African provinces', *South African Journal of Economics* **69**(2), 336–358.

[†] PROFESSOR AND HEAD, DEPARTMENT OF ECONOMICS, DIRECTOR, HEALTH AND DEVELOPMENT
POLICY RESEARCH GROUP, UNIVERSITY OF PRETORIA, PRETORIA, REPUBLIC OF SOUTH AFRICA;
(O) 27-12-420-5285, (F) 27-12-362-5207.

E-mail address: `steve.koch@up.ac.za`

[‡] GRADUATE STUDENT, DEPARTMENT OF ECONOMICS, UNIVERSITY OF PRETORIA, PRETORIA
0002, REPUBLIC OF SOUTH AFRICA.

E-mail address: `Jean.Slabbert@bafokeng.com`