Multi-Stage Stochastic Frontier Analysis for Simple Networks

Geraint Johnes[1], Mike Tsionas[1,2], and Marwan Izzeldin[1]

[1]Lancaster University Management School, LA1 4YX, U.K.
[2]Montpellier Business School
g.johnes@lancaster.ac.uk, m.tsionas@lancaster.ac.uk and
m.izzeldin@lancaster.ac.uk

July 21, 2024

Abstract

We develop a method for modelling multi-stage production using stochastic frontier analysis. This approach is suitable for the analysis of costs or output where intermediate outputs become inputs into a subsequent stage of the production process, either within an organization or in the form of a supply chain. Our focus is on higher education institutions in England, and the purpose is to assess the performance of our novel methods using MCMC methods. Without taking into full account of the complexity of the 'network', key decisions cannot be made regarding intake quality, student/staff ratios, per-student spending or academic reputation (the last of which involves costly decisions in terms of academic openings and the profile of candidates desired for any given university).

Keywords: Networks; Stochastic Frontier Analysis

JEL Classifications: C11, C13.

The authors acknowledge, without implication, extremely helpful comments from participants at the 2017 Budapest Workshop on Efficiency in Education, Jill Johnes, Ioannis Bournakis and Aya Ghalayini. Sadly, Mike Tsionas passed away in January 2024; the surviving authors dedicate this paper to his memory.

1 Introduction

Efficiency and overall performance of higher education is important for policy analysis as it is view that education is a production process where certain inputs are transformed into a single output can be challenged on two grounds. *First*, education is a multi-output process where degree results, employability and student satisfaction are produced based on inputs like research reputation, in-take quality, student/staff ratios, and per-student spending. In the existing literature, data envelopment analysis (DEA) has been extensively used to adopt such a multi-output approach where there are also multiple inputs (Thanassoulis et al., 2011; Lee and Johnes, 2022), but with only a few exceptions (Johnes, 2014) conventional statistical models based on stochastic frontier analysis have been confined to single output analyses of educational production (Johnes and Soo, 2017). *Second*, education is inherently a multi-stage process and the traditional view of simply transforming inputs into outputs has serious shortcomings. Although degree results arguably depend mostly on intake quality and resources such as the student/staff ratio, student satisfaction is determined by degree results (an intermediate output) and per-student spending while employability mostly depends on research reputation (which employers utilize as a signal in their decisions) and degree results, with per-student spending affecting employability only indirectly.Per-student spending is unobserved by employers who must base their assessment of a graduate's quality on the basis of degree results and institutional reputation. Our view of the education process is that, there are three stages or outputs: Stage A is degree results, stage B is employability (which is the ultimate goal of the teaching function in universities) and stage C which is student satisfaction (which affects future demand for education services from prospective students).

Ignoring the multi-stage nature of the education process has serious implications for performance. In the naive view where output is produced using a vector of inputs, we obtain an overall performance measure of efficiency and /or productivity. However, we do not know which stage of production is mostly responsible for this measure. Additionally, ignoring multi-stage production can result in serious misspecifications which directly invalidate overall performance measures. We argue that policymakers as well as universities are interested in identifying at which stage they perform worst (A, B or C?) rather than on how they perform overall. This question has been addressed using network DEA methods by Johnes (2013), but our aim here is to use a statistical approach to analyse the network. This offers the considerable advantage of estimating marginal effects which contain potentially useful information, as well as allowing the use of the full toolbox of statistical inference.

An overall performance indicator does not provide universities with a sense of how to allocate their limited resources effectively. It is of course true that universities target employability alongside other objectives but *the key question is how employability can be improved*. This question is by no means trivial as employability depends on degree results and reputation but degree results depend on inputs such as intake quality and student/staff ratio. What is the relative contribu-

tion of reputation versus degree results? What is the relative importance of reputation versus degree results when it comes to employability? What is the relative importance of intake quality versus student/staff ratios in determining degree results which, in turn, determine employability? Moreover, before embarking on specific actions regarding the inputs, are there any slacks or opportunities for improving performance in terms of degree results, employability, and student satisfaction?

Clearly, this is a complicated process that influences decision-making, but it reflects better than naive models the reality of provision of educational services. Without taking into full *account* the complexity of the 'network', key decisions cannot be made regarding intake quality, student/staff ratios, per-student spending, and academic reputation (the last of which involves costly decisions in terms of academic openings and the profile of candidates desired for any given university). Even so, the network that we consider may be considered a simplification - for example, we consider the impact of research reputation on the employability of graduates, but not on their degree results, since the latter are likely to be primarily determined by the teaching, rather than the research, function of the institution.

Once performance measures are derived at each node or stage of the education 'network', one can identify where slacks exist and these can be eliminated by focusing more on the particular nodes using relatively inexpensive (that is mostly administrative) policy actions. Effectively, this can result in cost savings relative to uninformed decisions made based on presumptions that more investment should be allocated to employability, reputation or per-student spending. Model-based shadow pricing allows us to monetize the cost of existing slacks and provide direct monetary measures of lack of good performance at the various nodes of the educational services provision. Although decision-making seems quite complicated, as we have argued, that statistical modelling is relatively straightforward to understand and implement, and provides a wealth of information that can be utilized by policymakers and decision-makers in universities.

Additionally, we provide a way to assess the relative importance of each input in each node of the process via marginal effects. These are non-trivial to compute because introducing slacks into an otherwise linear-in-the-parameters model converts it to a highly nonlinear model. Marginal effects can be used to facilitate the resource allocation and decision-making process, as they provide a more complete picture of how sensitive are different nodes or outputs to the various inputs. Unlike traditional models, where all inputs simultaneously produce all outputs (often represented by a transformation function such as an output- or input-oriented distance function, see, Coelli et al. (2005)), our approach allows for the determination of these marginal effects. This is particularly important in the context of educational services provision, where the traditional view of production does not facilitate such detailed analysis. Addressing this limitation of the traditional view in the application of SFA models is a key motivation for our paper.

The method of stochastic frontier analysis, developed by Aigner et al. (1977) has become widely used in contexts where production functions are to be estimated. The popularity of the

technique derives from the fact that it estimates the parameters of a production function as a frontier, recognising that observations may fall short of that frontier as a consequence of technical inefficiency. While, in many contexts, we might expect such inefficiency to be competed away, in situations where markets operate imperfectly it seems reasonable to expect that efficiency is distributed across producers in a non-degenerate way. The standard assumption underpinning these models is that inefficiency can be measured by exploiting an asymmetry in the error terms of a production function estimated by maximum likelihood, these being composed of two elements, namely inefficiency and the usual white noise.

In applied contexts, there is often scope for inefficiency to affect units of production at different points. An original equipment manufacturer (OEM) that sells the finished product to a client typically relies on a complex network of firms in its supply chain in order to produce its output and inefficiency can arise at any point within that network. Indeed, since the boundaries of the firm may be defined in different ways (Coase, 1937), it may often be the case that inefficiency can arise at different points, or nodes, of a network of economically significant connections within a single organization.

The analysis of efficiency in networks has hitherto relied on non-statistical methods. The linear programming approach of data envelopment analysis (DEA) has been used to construct networks within which efficiency may vary - and may be evaluated - at a variety of nodes (Färe (1991); Färe and Grosskopf (1996) ; Tone and Tsutsui (2009)). These techniques have been applied in a wide range of contexts, notably agricultural production (Rodríguez et al. (2014)), team sports (Moreno and Lozano (2014)) and supply chains (Lozano and Adenso-Diaz (2018)). An application of such methods in the context of higher education is provided by Johnes (2013). However, there are advantages to developing statistical alternatives to these non-parametric methods. In particular, the statistical approach of stochastic frontier analysis allows the full armoury of statistical inference to be employed.

Our aim in this paper is therefore to develop a stochastic frontier analysis model in which production is of a multi-stage character. In so doing, we adopt a Bayesian perspective, building on the work of Van den Broeck et al. (1994) which involves competition across a variety of assumptions about the form and parameters of the inefficiency distribution. Osiewalski and Steel (1998) presents a more general framework, and introduces into this context the MCMC approach of Gibbs sampling (Geman and Geman, 1987) as a means of making posterior inferences on the parameters of the production function and on the estimated efficiency scores. Our approach essentially generalises this from a single equation model to a network characterised by several equations representing a recursive system. In considering a system in which a multiplicity of inputs is used to produce more than one output, our approach resembles that of Fernandez et al. (2002, 2005), but the recursive nature of our system, designed to be analogous to the type of network analysed by Färe (1991) and others, is distinctive and necessitates innovation in our method.

We illustrate the operation of our method using data on producers of higher education. This

is a complex multi-product industry that exhibits the multi-stage properties characteristic of this type of problem in that the outputs of some constituent parts of a higher education producers become inputs for other parts. For example, the the output of a bachelor's program becomes an input into postgraduate programs, or - as here - the education process that leads to credentials is part of a larger system in which the ultimate aim is successful employment. There are thus several nodes within the multi-stage process that constitutes any single higher education institution and efficiency may vary across these nodes.

The rest of the paper is structured as follows. In section 2 we develop the econometric model. Section 3 provides the empirical example, in which efficiencies of higher educational institutions are evaluated. This is followed in section 4 by a more detailed evaluation of the model and its application using MCMC methods, and conclusions are drawn in section 5.

2 Econometric Model

2.1 Model Specification

In general form, we can write the equations for a multi-stage model as follows. Suppose we have M nodes or distinct production stages where (log) inputs $\mathbf{x}_{it,1}, \mathbf{x}_{it,2}, ..., \mathbf{x}_{it,M}$ are used, each one being a $K \times 1$ vector. Outputs are denoted by $Y_{it,1}, Y_{it,2}, ..., Y_{it,M}$ and $\theta \in \Theta \subseteq \Re^d$ is a parameter vector. A general multi-stage model, in which production is sequential, can be written as follows:

$$\begin{aligned} y_{it,1} &= \mu_{i1} + f_1(\mathbf{x}_{it,1}; \theta) + v_{it,1} - u_{it,1}, \\ y_{it,2} &= \mu_{i2} + f_2(y_{it,1}, \mathbf{x}_{it,2}; \theta) + v_{it,2} - u_{it,2}, \\ y_{it,3} &= \mu_{i3} + f_3(y_{it,1}, y_{it,2}, \mathbf{x}_{it,3}; \theta) + v_{it,3} - u_{it,3}, \\ &\cdots \\ y_{it,M} &= \mu_{iM} + f_M(y_{it,1}, y_{it,2}, ..., y_{it,M-1}, \mathbf{x}_{it,M}; \theta) + v_{it,M} - u_{it,M}. \end{aligned} \tag{1}$$

In this system $\mu_{i1}, ..., \mu_{iM}$ are individual effects, $y_{it,m} = \log Y_{it,m} - \log Y_{it,1}, m = 2, ..., M$, $y_{it,1} = \log Y_{it,1}$, $\mathbf{v}_{it} = [v_{it,1}, v_{it,2}, ..., v_{it,M}]'$ is a vector of two-sided error terms, $\mathbf{u}_{it} = [u_{it,1}, u_{it,2}, ..., u_{it,M}]'$ is a vector of one-sided error terms representing technical inefficiencies and $f_1(), f_2(), ..., f_M()$ are output distance functions that represent the technology in each stage. We note parenthetically that, for a general output oriented distance function of the form $D(Y_{it,1}, ..., Y_{it,M}, X_{it,1}, ..., X_{it,K}; \theta) = 1$, it is well known that we can exploit the homogeneity of degree one in outputs and use logs to express the distance function in the form $\log Y_{it,1} = f_1 (\log Y_{it,2} - \log Y_{it,1}, ..., \log Y_{it,M} - \log Y_{it,1}, \log X_{it,1}, ..., \log X_{it,K}; \theta)$. If production is not sequential but there is some jointness we can write the system as follows:

$$
\begin{aligned}
y_{it,1} &= \mu_{i1} + f_1(y_{it,2}, ..., y_{it,M}, \mathbf{x}_{it,1}; \theta) + v_{it,1} - u_{it,1}, \\
y_{it,2} &= \mu_{i2} + f_2(y_{it,1}, y_{it,3}, ..., y_{it,M}, \mathbf{x}_{it,2}; \theta) + v_{it,2} - u_{it,2}, \\
y_{it,3} &= \mu_{i3} + f_3(y_{it,1}, y_{it,2}, y_{it,4}, ..., y_{it,M}, \mathbf{x}_{it,3}; \theta) + v_{it,3} - u_{it,3}, \\
&\dots \\
y_{it,M} &= \mu_{iM} + f_M(y_{it,1}, y_{it,2}, ..., y_{it,M-1}, \mathbf{x}_{it,M}; \theta) + v_{it,M} - u_{it,M}.
\end{aligned} \tag{2}
$$

Some zero restrictions must be present unless there is full jointness in production. For example, if production in a given node uses as inputs the outputs from the two neighbouring nodes and the nodes can be arranged in a circle, we have:

$$
\begin{aligned}
y_{it,1} &= \mu_{i1} + f_1(y_{it,2}y_{it,M}, \mathbf{x}_{it,1}; \theta) + v_{it,1} - u_{it,1}, \\
y_{it,2} &= \mu_{i2} + f_2(y_{it,1}, y_{it,3}, \mathbf{x}_{it,2}; \theta) + v_{it,2} - u_{it,2}, \\
y_{it,3} &= \mu_{i3} + f_3(y_{it,2}, y_{it,4}, \mathbf{x}_{it,3}; \theta) + v_{it,3} - u_{it,3}, \\
y_{it,4} &= \mu_{i4} + f_3(y_{it,3}, y_{it,5}, \mathbf{x}_{it,4}; \theta) + v_{it,4} - u_{it,4}, \\
&\dots \\
y_{it,M} &= \mu_{iM} + f_M(y_{it,1}, y_{it,M-1}, \mathbf{x}_{it,M}; \theta) + v_{it,M} - u_{it,M}.
\end{aligned} \tag{3}
$$

To generalize (2) we can include the inputs from the other stages of the system:

$$
\begin{aligned}
y_{it,1} &= \mu_{i1} + f_1(y_{it,2}, ..., y_{it,M}, \mathbf{x}_{it,1}, ..., \mathbf{x}_{it,M}; \theta) + v_{it,1} - u_{it,1}, \\
y_{it,2} &= \mu_{i2} + f_2(y_{it,1}, y_{it,3}, ..., y_{it,M}, \mathbf{x}_{it,1}, ..., \mathbf{x}_{it,M}; \theta) + v_{it,2} - u_{it,2}, \\
y_{it,3} &= \mu_{i3} + f_3(y_{it,1}, y_{it,2}, y_{it,4}, ..., y_{it,M}, \mathbf{x}_{it,1}, ..., \mathbf{x}_{it,M}; \theta) + v_{it,3} - u_{it,3}, \\
&\dots \\
y_{it,M} &= \mu_{iM} + f_M(y_{it,1}, y_{it,2}, ..., y_{it,M-1}, \mathbf{x}_{it,1}, ..., \mathbf{x}_{it,M}; \theta) + v_{it,M} - u_{it,M}.
\end{aligned} \tag{4}
$$

While some zero restrictions must be imposed for identification, this depends on the particular multi-stage model under consideration. If the system in (4) were linear and $\tilde{K}_m$ inputs are excluded from the mth equation ($m = 1, ..., M$), the order condition for identification of the mth equation would be $\tilde{K}_m > M-1$. Therefore, if the number of inputs is large relative to the outputs, the order condition would be satisfied. Of course, if the system is nonlinear (for example, of the translog type) nonlinearity aids in identification and less stringent conditions would be required but an examination of the identification issue would require the examination of the specific structure of the particular multi-stage model.

It is useful at this stage to introduce the empirical application that will be explored later in the paper. We examine the production process within the tertiary schooling sector, where institutions of higher education (typically universities) uses a variety of inputs to produce graduates that are subsequently active in the labor market. The process of production can be viewed as a multi-stage model in which a set of inputs is used to produce an intermediate product that itself, along with further exogenous inputs, is subsequently used to produce the final output. This view of a system in which graduation represents an intermediate output in a chain where the final output

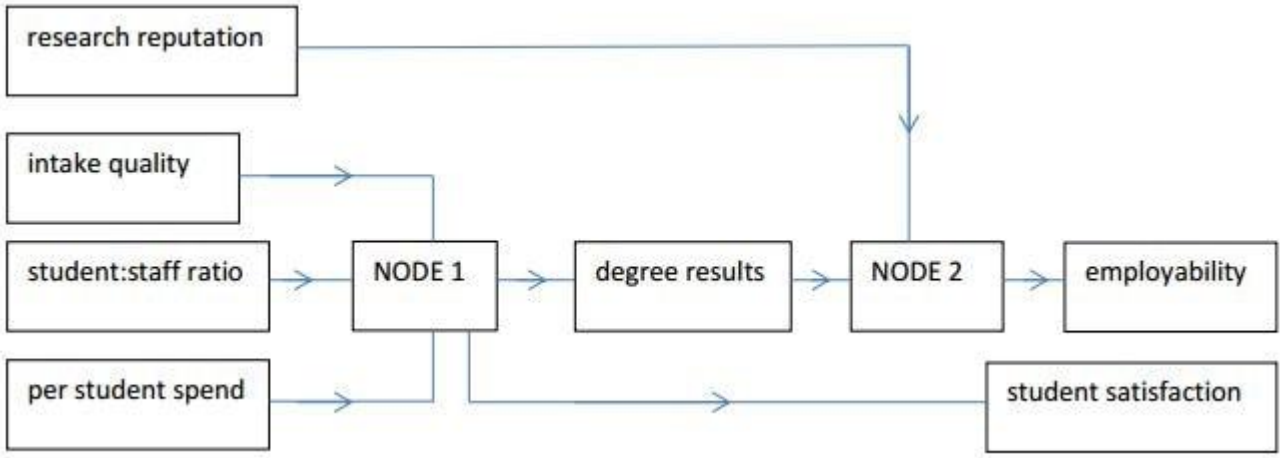

Figure 1: The Multi-Stage Process: Application to Education

is employability concords with the human capital view (Becker et al., 1964) in which education is undertaken as an investment, the ultimate goal of which is to secure an enhanced stream of future earnings (Suleman, 2018). Suppose the first set of inputs is $\mathbf{x}_1 = \begin{bmatrix} \text{intake quality} \\ \text{staff:student ratio} \\ \text{per student spend} \end{bmatrix}$, the second input is $x_2 =$ research reputation, and the outputs are $y_1 =$ degree results, y_2 is employability, and y_3 is student satisfaction. See Figure 1 for an illustration of this multi-stage process. This specification draws heavily on that used by Johnes (2013), and motivation for this model structure is provided in that paper.

Hence, for this specific example, equations (4) may be rewritten as

$$Y_{it,1} = \mu_{i1} + f_1(\mathbf{X}_{it,1}; \theta) + v_{it,1} - u_{it,1}, \tag{5}$$

$$Y_{it,2} = \mu_{i2} + f_2(\mathbf{X}_{it,1}, \mathbf{X}_{it,2}, Y_{it,1}; \theta) + v_{it,2} - u_{it,2}, \tag{6}$$

$$Y_{it,3} = \mu_{i3} + f_3(\mathbf{X}_{it,1}; \theta) + v_{it,3} - u_{it,3}, \; i = 1, ..., n, \; t = 1, .., T, \tag{7}$$

where capital letters are used for inputs and outputs. The inputs $\mathbf{X}_{it,1}$ in (6) are included for generality; whether the weights associated with them are non-zero is an empirical issue. Our statistical assumptions are:

$$\mathbf{v}_{it} = [v_{it,1}, v_{it,2}, v_{it,3}]' \sim \mathcal{N}(0, \Sigma), \tag{8}$$

$$\mathbf{u}_{it} = [u_{it,1}, u_{it,2}, u_{it,3}]' \sim \mathcal{N}_+(0, \Omega), \tag{9}$$

$$\mu_{ij} \sim \mathcal{N}(0, \sigma^2_{\mu j}), j = 1, 2, 3, \tag{10}$$

though of course other assumptions for the distributions of the efficiencies - such as the exponential - are in principle possible. A multivariate exponential distribution may be implemented as follows. Noting that, in the univariate case, the probability density function of the nonnegative random variable $\mathbb{U}$ is: $f(u) = \alpha \exp(-\alpha u)$, $u \geq 0$, where $\alpha > 0$, the extension to the multivariate case implies:

$$u_{tm} = \omega_{to} + \omega_{tm}, \ \forall m = 1, ..., M, \tag{11}$$

where ω_{to} follows an exponential distribution with parameter α_0 and ω_{tm} follow, independently, an exponential distribution with parameter α_m $(m = 1, ..., M)$. Moreover, ω_{to} and $\{\omega_{tm}, m = 1, ..., M\}$ are independent. The u_{tm}s are correlated through the common component ω_{to}. For this construction, see Kotz et al. (2000, pp. 454-455) which they attribute to Cherian (1941) and Ramabhadran (1951).

The first two equations essentially define a recursive system so the Jacobian of the system in equations (5)-(7) is unity. It is, of course, possible to adopt an agnostic distance function framework into which all inputs are used to produce simultaneously the three outputs. The drawbacks of this approach are that (i) it does not account for the sequential character of production, and (ii) it can only identify an overall or catch-all inefficiency term whereas, in fact, there are three sources of inefficiency in the system. In our framework we allow for correlated inefficiencies in equation (9). All our functional forms in (5)-(7) are given by the translog specification which we write generically as follows:

$$f(x_{it}; \theta) = a_o + a' x_{it} + \frac{1}{2} x'_{it} A x_{it}, \tag{12}$$

where a_0, a and A are vectors and a symmetric matrix of parameters respectively.

The Jacobian matrix of transformation from $\mathbf{v}_{it}$ to $\mathbf{y}_{it} = [y_{it,1}, ..., y_{it,M}]'$ is given by the following general expression:

$$\mathscr{J}_{it}(\theta) = \begin{bmatrix} 1 & g_{it,12} & \cdots & g_{it,1M} \\ g_{it,21} & 1 & \cdots & g_{it,2M} \\ \vdots & \vdots & \vdots & \vdots \\ g_{it,M1} & g_{it,M2} & \cdots & 1 \end{bmatrix}, \tag{13}$$

where $g_{it,mm'} = -\frac{\partial f_m}{\partial y_{it,m'}}$ $(m \neq m', m, m' = 1, ..., M)$. In the log-likelihood function and the log posterior density the term that is introduced by the transformation is $\sum_{i=1}^{n} \sum_{t=1}^{T} \log |\det \mathscr{J}_{it}(\theta)|$. Although this term depends on the data, posterior analysis by Markov Chain Monte Carlo (MCMC) methods should usually be feasible.

2.2 Bayesian Analysis

In the construction of the likelihood, we assume $\mathbf{v}_{it}$, $\mathbf{u}_{it}$ and μ_{ij} are uncorrelated and further the μ_{ij}s are uncorrelated with the explanatory variables. The likelihood function of the model in (5)-(7) under the stochastic specification assumptions in (8)-(9) can be written as follows:

$$L(\theta,\Sigma,\Omega,\{\sigma_j^2\};\mathcal{Y}) \propto |\Sigma|^{-nT/2}|\Omega|^{-nT/2}\int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\mathbf{V}_{it}(\theta)+\mathbf{u}_{it}\right]'\Sigma^{-1}\left[\mathbf{V}_{it}(\theta)+\mathbf{u}_{it}\right]\right\}\cdot C(\Omega)^{nT}\cdot|\Omega|^{-(nT+\underline{\nu}+1)/2}\cdot\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{u}_{it}'\Omega^{-1}\mathbf{u}_{it}\right\}\cdot\exp\left\{\sum_{i=1}^{n}\sum_{t=1}^{T}\log|\det\mathscr{J}_{it}(\theta)|\right\}\cdot\left[\prod_{j=1}^{M}\sigma_{\mu j}^{-n}\exp\left\{-\frac{\sum_{i=1}^{n}\mu_{ij}^2}{2\sigma_{\mu j}^2}\right\}\right]d\mathbf{u}d\boldsymbol{\mu}, \tag{14}$$

where $\mathcal{Y}$ denotes the entire data set,
$\mathbf{V}_{it}(\theta)=[y_{it,1}-f_1(\mathbf{x}_{it,1};\theta)-\mu_{i1}, y_{it,2}-f_2(\mathbf{x}_{it,1},x_{it,2},y_{it,1};\theta)-\mu_{i2},...,y_{it,M}-f_M(\mathbf{x}_{it,1};\theta)-\mu_{iM}]'$,$\boldsymbol{\mu}=[\mu_{ij},i=1,...,n,j=1,...,M]$, $\mathbf{u}=[u_{it}]$, and $C(\Omega)=(2\pi)^{-k/2}\int_A\exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{z}\right\}d\mathbf{z}$, where the range of integration A, is defined by the inequalities $P^{-1}\mathbf{z}\geq\mathbf{0}_k$, where $M=3$ in our case, and $\Omega^{-1}=P'P$, where P is lower triangular. Moreover the vector of all $\{\mu_{im},m=1,...,M\}$ is denoted by $\boldsymbol{\mu}$. We can write (14) as follows:

$$L(\theta,\Sigma,\Omega,\{\sigma_m^2\}_{m=1}^{M};\mathcal{Y}) \propto |\Sigma|^{-nT/2}|\Omega|^{-nT/2}\int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}}\exp\left\{-\tfrac{1}{2}\mathrm{tr}\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})\Sigma^{-1}\right\}\cdot C(\Omega)\cdot\exp\left\{-\tfrac{1}{2}\mathrm{tr}\mathbf{B}(\theta,\mathbf{u})\Omega^{-1}\right\}\cdot\prod_{j=1}^{M}\left[\sigma_{\mu j}^{-n}\exp\left\{-\frac{\sum_{i=1}^{n}\mu_{ij}^2}{2\sigma_j^2}\right\}\right]d\mathbf{u}d\boldsymbol{\mu}, \tag{15}$$

We use the following result: $\sum_{i=1}^{N}v_i'\Sigma^{-1}v_i=trA\Sigma^{-1}=tr\Sigma^{-1}A$,

where v_i is $N\times K$ (K denotes the number of rows or equations and v_is are 'residuals'), $A=\sum_{i=1}^{N}v_iv_i'$ is $K\times K$ and Σ is $K\times K$.

where $\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})=\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\mathbf{V}_{it}(\theta)+\mathbf{u}_{it}\right]\left[\mathbf{V}_{it}(\theta)+\mathbf{u}_{it}\right]'$, $\mathbf{B}(\theta,\mathbf{u})=\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{u}_{it}\mathbf{u}_{it}'$. Given a prior $p(\theta)$ Bayes' theorem yields the posterior:

$$p(\theta|\mathcal{Y})\propto L(\theta;\mathcal{Y})p(\theta). \tag{16}$$

Our prior, $p(\theta)$ is flat over the domain,$\mathcal{R}$, ensuring monotonicity and concavity of all translog functional forms (see, Gallant and Golub (1984)). Therefore:

$$p(\theta)\propto \text{const.}\mathbb{I}(\theta\in\mathcal{R}), \tag{17}$$

For Σ and Ω we use priors of the form:

$$p(\Sigma)\propto|\Sigma|^{-(\underline{\nu}+1)/2}\exp\left\{-\underline{A}\Sigma^{-1}\right\},\ p(\Omega)\propto|\Omega|^{-(\underline{\nu}+1)/2}\exp\left\{-\underline{A}\Omega^{-1}\right\}, \tag{18}$$

where $\underline{\nu}=1$ and $\underline{A}=10^{-4}I$. These are in the Wishart family and they are relatively diffuse,

see Zellner (1971a), page 395. The zero mean assumption on the random effects is justified as we always include an intercept. For the random effect variances (see Zellner (1971a), page 371) our prior is,

$$\frac{\underline{q}}{\sigma_{\mu j}^2} \sim \chi^2(\underline{\nu}),\ j = 1, 2, 3. \tag{19}$$

Finally, regarding (19) we set $\underline{\nu} = 0$ and $\underline{q} = 10^{-4}$ so that these priors are, practically, flat.

The monotonicity and concavity restrictions are ensured using rejection sampling as the number of observations is approximately only 200. We are satisfied when approximately 95% of all observations satisfy the monotonicity and concavity restrictions and that our translog function is well-behaved. The average number of rejections per MCMC iteration and observation was 5. We have explored the issue of regularity conditions and are satisfied that we find enough regions in the input-output space where the restrictions of monotonicity and concavity are met. Specifically, we have imposed regularity at the mean vector of all variables +/- h times the vector of standard deviations where h=2 initially, and then checked that the conditions hold at other points. Technical efficiency can be computed easily as:

$$r_{it,j} = S^{-1} \sum_{s=1}^{S} \exp\left\{-u_{it,j}^{(s)}\right\},\ i = 1, ..., n,\ t = 1, ..., T, \tag{20}$$

for the s-th draw of the Gibbs sampler. To implement the Gibbs sampler we use 10,000 draws preceded by 5,000 to mitigate the possible impact of start-up effects and ensure convergence (Geweke, 1992). Details for our MCMC procedures are provided in Appendix 1. Additionally, we use a Riemannian MCMC procedure due to Girolami and Calderhead (2011). The algorithm uses first and second derivative information from the log posterior and seems to perform particularly well. We implement the procedure using 15,000 passes the first 5,000 of which are discarded in the burn-in phase. The two techniques yield virtually the same results but the Riemannian MCMC procedure behaves better in terms of autocorrelation. Therefore, we decided to report results based on the Riemannian MCMC procedure but also make sure that we have a Gibbs sampler (the first of our MCMC techniques) for comparison purposes and ensuring convergence to the correct posterior distribution of the parameters and latent variables.

3 Empirical Example

The data used in the analysis that follows refer to higher education institutions in England, and form a short balanced panel over a recent two-year period. The removal of an institutional cap on the number of students that could be recruited by each university around the period to which our data refer limits in practice the extent of correlation between input measures, and between these and the fixed effects. Most of the data come from the Guardian University Rankings for

2016 and 2017, these rankings are based on original data attached to earlier academic years. Data on research performance are drawn from the Research Excellence Framework (REF) grade point averages published at Times Higher Education. The 2014 REF assessed research performance over the 2008-13 period. In contrast to the other measures used here, this does not vary over the two years in the panel. Data on degree results - specifically the percentage of undergraduates completing with good degrees (classified as first or upper-second class) - come from the Higher Education Statistics Agency publication, Students in Higher Education Institutions. The institutions included in the sample are all traditional institutions in the English context; âalternative providersâ, a new generation of institutions that are privately owned, that may be for-profit, and that are typically small are not included.

The model estimated here is as given in (6) through (8) with the μ modelled as random effects owing to the short nature of the panel. Descriptive statistics for efficiency are in Table 1, both under half-normal and multivariate exponential cases, and the sample distributions of posterior mean efficiencies (see equation (20)) are reported in Figure 2. While the availability of efficiency scores at each node obviates the need for an aggregate score, such that other authors (for example, Fernandez et al. (2002)) do not report an aggregate, in the interests of completeness we effect an aggregation as follows. Denoting output prices by p_1, p_2, p_3 and input prices by $\mathbf{w}_1$ and w_2, note that profit maximization implies:

$$\begin{aligned} p_1 \frac{\partial f_1}{\partial \mathbf{x}_1} + p_2 \left(\frac{\partial f_2}{\partial \mathbf{x}_1} + \frac{\partial f_2}{\partial y_1} \frac{\partial f_1}{\partial \mathbf{x}_1} \right) + (1 - p_1 - p_2) \frac{\partial f_3}{\partial \mathbf{x}_1} = \mathbf{w}_1, \\ p_2 \frac{\partial f_2}{\partial x_2} = w_2. \end{aligned} \tag{21}$$

Normalizing so that $p_1 + p_2 + p_3 = 1$ the above system can be solved to obtain p_1, p_2 and p_3 given values for the input prices.

The problem with this approach is that input prices are not available. In this work, we propose to model input prices using latent dynamic processes, viz.:

$$\mathbf{w}_{it} = \begin{bmatrix} \mathbf{w}_{it,1} \\ w_{it,2} \end{bmatrix} = \mathbf{a}_i + \mathbf{A}\mathbf{w}_{i,t-1} + \boldsymbol{\zeta}_{it}, \tag{22}$$

where $\mathbf{w}_{it,1}$ represents the (i,t) observation on $\mathbf{w}_1$, $w_{it,2}$ represents the (i,t) observation on w_2, $\mathbf{a}_i$ represents individual effects, $\mathbf{A}$ is a matrix of coefficients, and $\boldsymbol{\zeta}_{it} \sim \mathcal{N}(0, \Sigma_\zeta)$ is a vector error term. Define $\frac{\partial f_{it,1}}{\partial \mathbf{x}_{it,1}} = D_{it,1}(\theta; \mathcal{Y})$ $\frac{\partial f i_{it,2}}{\partial \mathbf{x}_{it,1}} + \frac{\partial f_{it,2}}{\partial y_{it,1}} \frac{\partial f_{it,1}}{\partial \mathbf{x}_{it,1}} = D_{it,2}(\theta; \mathcal{Y})$, $\frac{\partial f_{it,3}}{\partial \mathbf{x}_{it,1}} = D_{it,3}(\theta; \mathcal{Y})$, and $\frac{\partial f_{it,2}}{\partial x_{it,2}} = D_{it,4}(\theta; \mathcal{Y})$. In turn, we can write (21) as follows:

$$\begin{bmatrix} p_{i1} D_{it,1}(\theta; \mathcal{Y}) + p_{i2} D_{it,2}(\theta; \mathcal{Y}) + p_{i3} D_{it,3}(\theta; \mathcal{Y}) \\ p_{i2} D_{it,4}(\theta; \mathcal{Y}) \end{bmatrix} = \mathbf{w}_{it}, \tag{23}$$

along with the specification in (22). In this system, we allow output prices to be different for

different universities but time-invariant. The steady-state for input prices resulting from (22), is:

$$\mathbf{w}_i^* = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{a}_i. \tag{24}$$

Given a prior on the elements of $\mathbf{A}$, and a prior for steady-state input prices, we can recover a prior for the individual effects which support the prior steady state input prices. (Based on a simple regression of costs against intake quality, student: staff ratio and per-student spend, we assume that these take the values 90, 90000 and 1000 respectively, and we assume a value of unity for the steady state prior to the price of research reputation.) Specifically, we use:

$$\mathbf{a}_i|\mathbf{w}_i^*, \mathbf{A}, \Sigma_\zeta, \Sigma_{w^*} \sim \mathcal{N}\left((\mathbf{I} - \mathbf{A})\mathbf{w}_i^*,\ h\left[(\mathbf{I} - \mathbf{A})\Sigma_{w^*}(\mathbf{I} - \mathbf{A})' + \Sigma_\zeta\right]\right), \tag{25}$$

where $h > 0$ is a parameter, $[(\mathbf{I} - \mathbf{A})\Sigma_{w^*}(\mathbf{I} - \mathbf{A})' + \Sigma_\zeta]$ is the covariance of individual effects resulting from the steady state of (22), and Σ_{w^*} represents the covariance matrix of steady-state input prices which we determine as follows. We use $h = 100$ to make this prior vague relative to the information provided by the term in brackets in (25) implying that we only impose information resulting from the steady state, in a soft manner. Moreover, we enforce stationarity by assuming that all eigenvalues of $\mathbf{A}$ have a modulus less than unity. For Σ_ζ we use the same prior as in (18).

Our strategy is to embed (23) and (22) in (16). Given MCMC draws $\left\{\theta^{(s)}, s = 1, ..., S\right\}$ for θ, $D_{it,j}(\theta; \mathcal{Y})$, $j = 1, ..., 4$ can be computed. For each draw, we run a Sequential Monte Carlo (SMC) procedure (see Appendix C) to provide estimates of the states $\mathbf{w}_{it}^*$ along with a Girolami and Calderhead (2011) procedure to provide draws for $\mathbf{A}, \Sigma_\zeta$, and $p_{ij}, j = 1, 2, 3$. In turn, we have university-specific output prices as well as university-specific input prices which provides a wealth of information relevant to the most important decisions by policymakers and university administrators. An additional advantage is that because of the presence of $\boldsymbol{\zeta}_{it}$, the profit maximization conditions in (21) need not hold exactly, so allocative inefficiency is allowed. This is particularly important as markets have limited importance in university decision making so, examining deviations from these profit maximization first order conditions is another criterion to evaluate the optimality of actual university and administration decisions.

This allows us to report the distribution of combined posterior mean efficiencies in Figure 2. As is readily observed, the mean efficiency scores at each node is high. In the case of the half-normal efficiency distribution, the mean achieved at stage B - representing the employability goal - is lower than that achieved for both degree results (an intermediate output) and student satisfaction. This being the case, it is unsurprising to note that the variability of performance across institutions is greater for the employability goal than for the other outputs. Indeed, two universities achieve efficiency scores lower than 0.8 in each year for this goal. This finding echoes the

	Multivariate Half-Normal				Multivariate Exponential			
	stage A	stage B	stage C	overall	stage A	stage B	stage C	overall
post. mean	0.935	0.905	0.952	0.932	0.931	0.942	0.985	0.935
post. median	0.944	0.921	0.958	0.922	0.939	0.958	0.991	0.926
post s.d.	0.031	0.056	0.025	0.087	0.031	0.057	0.025	0.092
5%	0.848	0.748	0.887	0.781	0.843	0.785	0.920	0.802
95%	0.975	0.974	0.983	0.970	0.970	0.997	0.999	0.999

Table 1: Sample statistics of posterior mean efficiencies

results obtained by Johnes (2013) using data for a (single) earlier year and using a non-parametric network DEA approach, and suggests that institutions may need to devote attention to increasing stage B efficiency - possibly by addressing the performance of their graduate placement and careers services. Indeed the broad similarity in results between the earlier network DEA work and the current study serves to reinforce confidence in both approaches. The correlation between efficiencies at each of the three nodes is low. Indeed, none of the off-diagonal elements of the correlation matrix of efficiencies exceeds 0.125.

The marginal effects on output (be it degree results, employability, or student satisfaction) of each of the explanatory variables, evaluated at mean values, are reported in Table 2 - again both for the half-normal and multivariate exponential cases. With the exception of the impact of per-student spending on student satisfaction, these are all significantly positive. The magnitudes of the marginal effects are instructive. Particularly noteworthy - and of interest to managers of higher education institutions - is the powerful effect of intake quality, and the relatively small effect of per-student spending, on student satisfaction. Note that the student-staff ratio variable appears here in inverse form. All of the marginal effects have the expected sign. An additional explanatory variable, namely a binary variable indicating that the institution gained university status in or after 1992, was included in the empirical counterpart of each of the equations (1), (2) and (3) in some of our early experimentation, but this proved to be insignificant in all equations. The 1992 date is notable for being the year in which many former polytechnics gained university status.

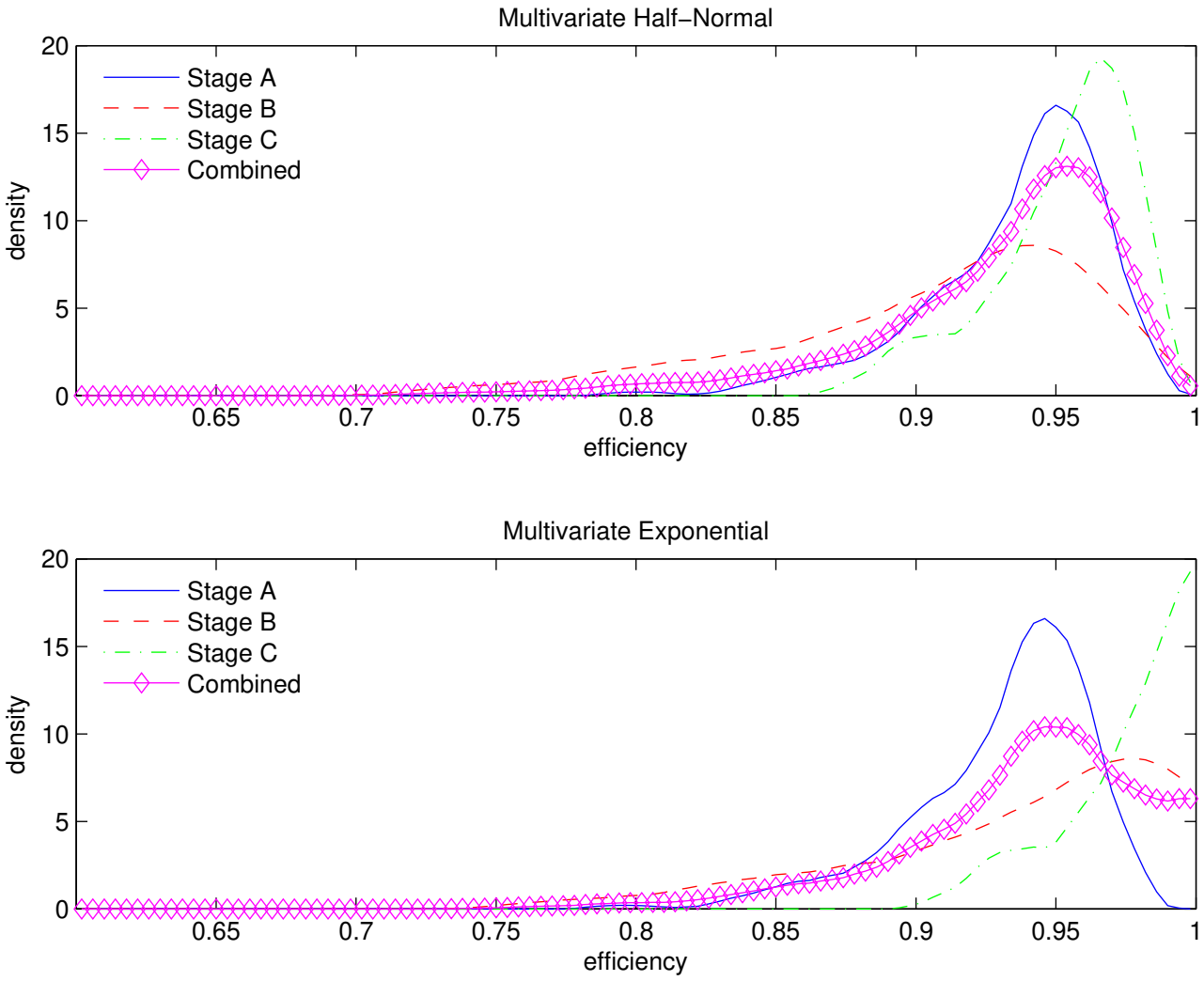

Figure 2: Sample distributions of posterior mean efficiencies

	Multivariate Half-Normal			Multivariate Exponential		
	stage A	stage B	stage C	stage A	stage B	stage C
intake quality	0.487	0.381	0.672	0.353	0.377	0.6871
	(0.0421)	(0.073)	(0.234)	(0.0388)	(0.069)	(0.215)
staff:student ratio	0.382	0.284	0.385	0.385	0.277	0.371
	(0.047)	(0.058)	(0.035)	(0.044)	(0.051)	(0.029)
per student spend	0.244	0.256	0.005	0.232	0.271	0.004
	(0.022)	(0.026)	(0.015)	(0.020)	(0.021)	(0.019)
x_2, research reputation	-	0.181	-	-	0.192	-
		(0.066)			(0.071)	
y_1, degree results	-	0.165	-	-	0.160	-
		(0.081)			(0.077)	
σ_v	0.060	0.053	0.038			
	(0.009)	(0.010)	(0.007)			
σ_u	0.083	0.128	0.062			
	(0.019)	(0.017)	(0.013)			
α_m				1.324	1.210	1.410
				(0.035)	(0.017)	(0.033)
α_0				0.988 (0.032)		

Notes: Standard deviation in parentheses. According to the posterior distributions, there is a 95% probability that these coefficients are greater than zero.

Table 2: Marginal Effects

The Bayesian approach adopted here provides a convenient method of estimating the parameters of the model in light of the integral in the likelihood. An alternative, though more computationally expensive, approach is to use maximum simulated likelihood. This approach is detailed in Appendix 2, where empirical results that correspond to those of Tables 1 and 2 are also reported. Results obtained using the two methods are congruent.

4 Monte Carlo Experiment

To examine the properties of our novel stochastic frontier model we design a Monte Carlo experiment to evaluate the model in equations (5)-(7). Posterior means are taken as the true values of θ. We assume $v_{it,m} \sim \mathcal{N}(0, \sigma_v^2)$ and $u_{it,m} \sim \mathcal{N}(0, \sigma_u^2)$ where the parameters σ_v and σ_u are, for simplicity, assumed common. The reparametrization $\lambda = \frac{\sigma_u}{\sigma_v}$ and $\sigma^2 = \sigma_v^2 + \sigma_u^2$ is adopted where λ denotes the signal-to-noise ratio. We set $\sigma = 0.5$ and a range of different values of λ is considered. The focus of interest is in the rank correlation between actual and estimated inefficiencies $u_{it,m}$ $(m = 1, 2, 3)$. In all cases, 15000 MCMC draws are used, the first 5,000 of which are discarded to mitigate possible start-up effects.

Results from the Monte Carlo experiment are reported in Table 3. It is clear from this table that the performance of the new method developed in this paper is acceptable in samples as small

		$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$
$n = 100$						
	$T = 5$	0.33	0.45	0.55	0.67	0.71
	$T = 10$	0.37	0.49	0.60	0.72	0.77
	$T = 25$	0.41	0.53	0.65	0.76	0.79
	$T = 50$	0.55	0.61	0.69	0.79	0.83
$n = 500$						
	$T = 5$	0.43	0.57	0.66	0.79	0.90
	$T = 10$	0.57	0.63	0.70	0.81	0.85
	$T = 25$	0.62	0.67	0.74	0.84	0.86
	$T = 50$	0.66	0.69	0.77	0.86	0.88
$n = 1,000$						
	$T = 5$	0.58	0.63	0.72	0.82	0.87
	$T = 10$	0.61	0.65	0.74	0.85	0.89
	$T = 25$	0.63	0.67	0.76	0.87	0.90
	$T = 50$	0.65	0.69	0.78	0.89	0.92
$n = 5,000$						
	$T = 5$	0.73	0.77	0.80	0.92	0.94
	$T = 10$	0.76	0.79	0.83	0.94	0.95
	$T = 25$	0.78	0.81	0.83	0.96	0.97
	$T = 50$	0.80	0.84	0.85	0.97	0.98

Table 3: Monte Carlo Results

as $n = 100$ and, asymptotically, the rank correlations between actual and estimated inefficiencies are quite large, especially when the signal-to-noise ratio exceeds 1. With $n = 5,000$ the rank correlations range from 0.80 to 0.85 when $\lambda = 1$ depending on the different values of T, and from 0.92 to 0.97 when $\lambda = 2$. In small samples ($n = 100$) the rank correlations range from 0.55 to 0.69 when $\lambda = 1$, 0.67 to 0.79 when $\lambda = 2$ and 0.71 to 0.83 when $\lambda = 5$.

‘

Since we use MCMC, it is essential to examine the convergence and autocorrelation of the MCMC draws. Convergence is assessed using Geweke (1992) convergence diagnostic. Specifically, for each parameter, we test for equality of means in the first 50% and last 25% of the draws. Geweke (1992) statistic converges asymptotically (in the number of draws) to a standard normal distribution. These statistics, in our application, range from 0.813 to 1.572 in absolute value, confirming that MCMC has converged. To examine autocorrelation, we compute autocorrelation functions (acf) for each parameter based on the MCMC draws. In Figure 3, we report the *maximal* values of autocorrelation coefficients at each lag from 1 to 50 (in absolute value but retaining the sign for plotting). For comparison, we also plot the *maximal* values of autocorrelation coefficients at each lag from a Metropolis-Hastings MCMC scheme: The Metropolis-Hastings MCMC scheme generates a candidate draw as $\theta^c \sim \mathcal{N}\left(\theta^{(s-1)}, hI\right)$, where $\theta^{(s-1)}$ is the previous draw and $h > 0$

is a smoothing constant. The candidate is accepted with probability $\min\left\{1, \frac{p(\boldsymbol{\theta}^c|\mathcal{Y})}{p(\boldsymbol{\theta}^{(s-1)}|\mathcal{Y})}\right\}$ and we set $\theta^{(s)} = \theta^c$, otherwise we repeat the previous draw and we set $\theta^{(s)} = \theta^{(s-1)}$, $\forall s = 1, ..., S$. We use the same number of draws, S, and we select h by trial-and-error so that approximately 25% of all candidates are, eventually, accepted. The Metropolis-Hastings MCMC scheme is applied to all parameters except Σ which can be integrated out analytically (Zellner, 1971b, p. 243, formula (8.86)).

From the autocorrelation functions in Figure 3, the performance of MCMC is seen to be much better in comparison with the Metropolis-Hastings MCMC as autocorrelations are practically zero after about lag 5 in the former case. By way of contrast, autocorrelations from Metropolis-Hastings MCMC are still close to 0.5 at lag 50.

A further interesting technical question is whether results are sensitive to the number of draws. To this purpose, we increase the number of draws to 25,000, 35,000,...,125,000. We omit the first 5,000 draws to mitigate possible start-up effects and we recompute posterior means and posterior standard deviations (s.d.) of parameters. The percentage deviations relative to the case with 15,000 draws are presented in Figure 4 from which it is evident that increasing substantially the number of draws produces virtually no change in results.

It would be interesting, however, to adopt an informative prior of the form:

$$\theta \sim \mathcal{N}\left(\underline{\theta}, \underline{\mathbf{V}}\right), \tag{26}$$

where $\underline{\theta}$, $\underline{\mathbf{V}}$ are, respectively, the prior mean and prior covariance matrix. This prior is truncated to the set where the production functions are monotone, increasing and concave to account for the economic restrictions. To examine sensitivity with respect to the prior, we set $\underline{\theta} = \varphi I$ and $\underline{\mathbf{V}} = \omega^2 I$. We generate 10,000 different φ and ω values from uniform distributions in the interval $(-10^6, 10^6)$ and $(1, 10^6)$ respectively. In turn, we rerun our MCMC scheme with 15,000 draws, the first 5,000 are discarded to mitigate start-up effects. We compute the percentage deviations of posterior means and posterior s.d. relative to the baseline case which corresponds to the prior we selected when we performed the initial computations. The results are reported in Figure 5, where we present kernel densities of the percentage deviations of posterior means and posterior s.d. relative to the baseline case. The kernel densities are computed using all parameter draws and all 10,000 different φ and ω values relative to the baseline case. The results reported in Figure 5 indicate robustness with respect to changes in the prior.

5 Conclusions

The development in this paper of a method for analysing a multi-stage stochastic frontier model provides a powerful tool. Where much production takes place in networks - whether these take the form of a supply chain or simply a series of connections between departments or functions of

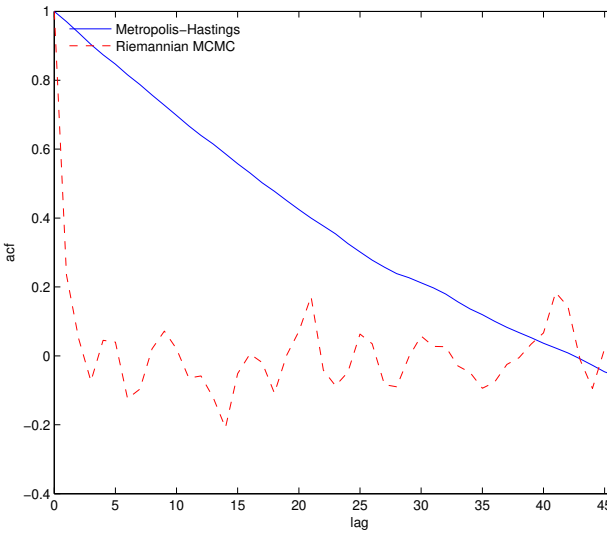

Figure 3: Autocorrelation functions

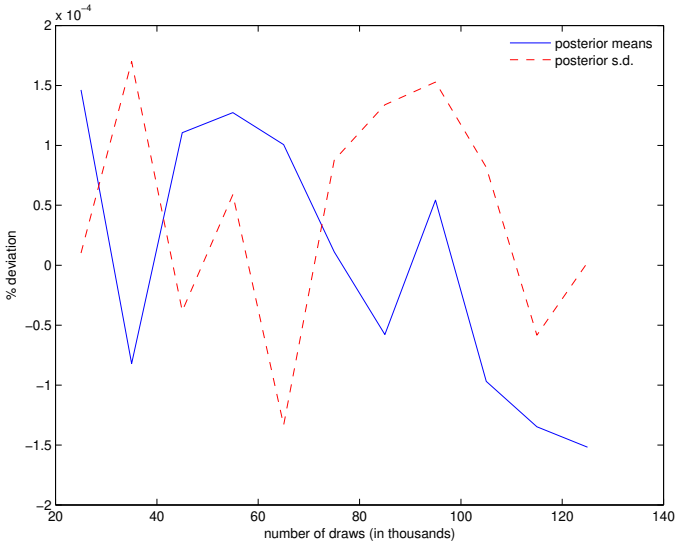

Figure 4: Percentage deviations of posterior means and s.d. from a different number of draws

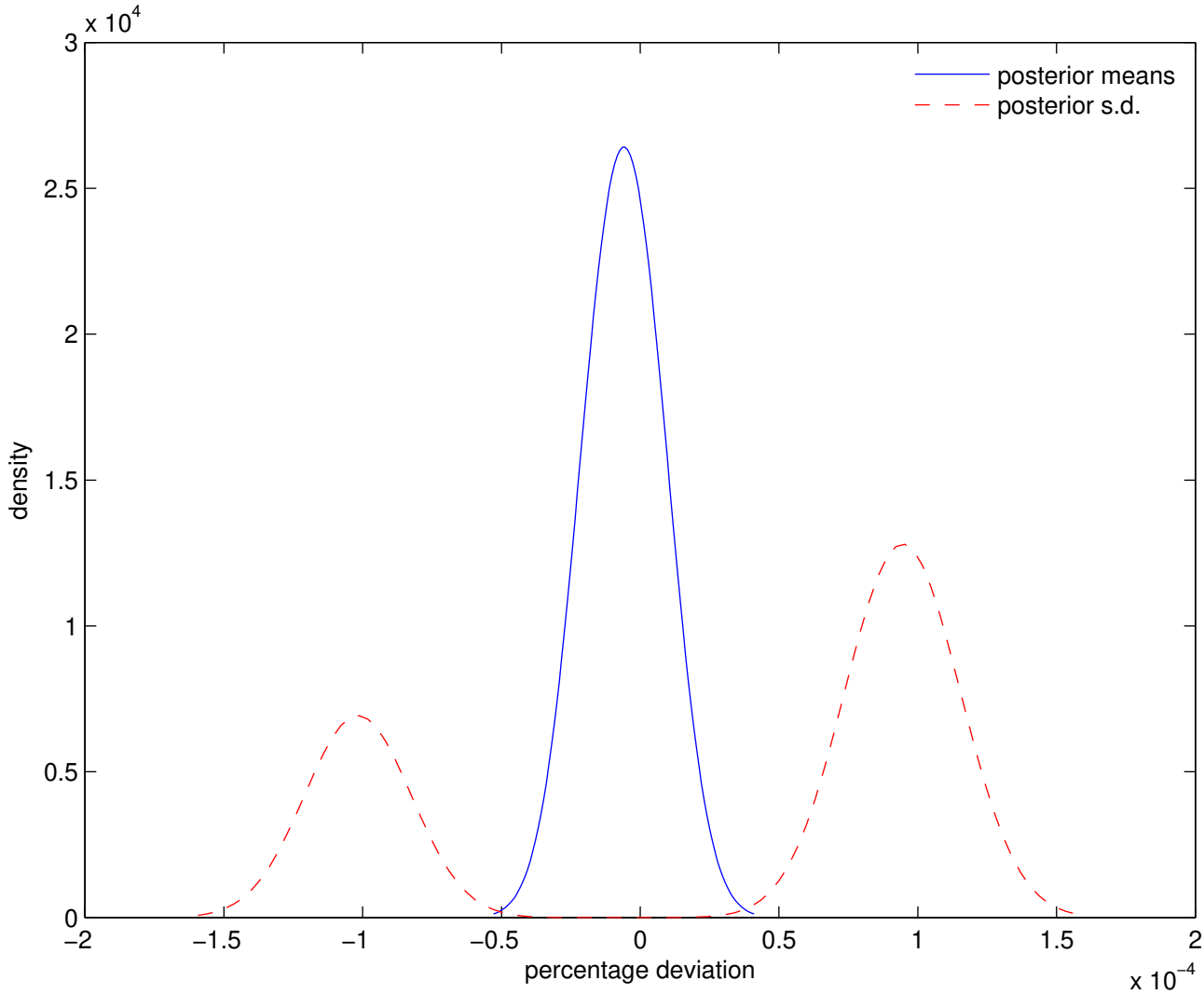

Figure 5: Percentage deviations of posterior means and s.d. from different priors

a single producer - it is useful to develop a model of production that explains the output at each node, and which is capable also of identifying efficiency, as a measure of performance, at each node. The approach promises to deliver a considerably enhanced understanding of what goes on inside the 'black box' that is conventionally summarized by a production function.

We have illustrated the method using data for universities in England. In line with earlier work, the analysis confirms that the greatest efficiency challenge faced by these institutions is attached to their employability function. Faced with the heightened prominence of rankings, universities have in recent years increased their investment in students' career development (Smith et al. (2000); Mason et al. (2009)), but a small number of institutions continue to perform at a level some distance short of the frontier in this dimension.

Clearly, the network considered here is a fairly simple one involving a chain of inputs, intermediate outputs and final outputs, and there are various ways in which it could be reconfigured. The methodology, however, can be routinely amended (changing equations 1 through 3) to represent considerably more complex characteristics of, for example, the supply chains observed in high-technology manufacturing sectors. Applications in such contexts promise to throw much new light on issues of productivity and company performance.

References

Aigner, D., Lovell, C.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Andrieu, C., Roberts, G.O., 2009. The pseudo-marginal approach for efficient monte carlo computations. The Annals of Statistics 37, 697–725.

Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behavioral science 9, 226–232.

Van den Broeck, J., Koop, G., Osiewalski, J., Steel, M.F., 1994. Stochastic frontier models: A bayesian perspective. Journal of Econometrics 61, 273–303.

Casarin, R., Marin, J., 2007. Online data processing: Comparison of Bayesian regularized particle filters. University of Brescia, Department of Economics. Technical Report. Working Paper.

Cherian, K., 1941. A bi-variate correlated gamma-type distribution function. The Journal of the Indian Mathematical Society 5, 133–144.

Coase, R.H., 1937. The nature of the firm. Economica 4, 386–405.

Coelli, T.J., Rao, D.S.P., O'donnell, C.J., Battese, G.E., 2005. An introduction to efficiency and productivity analysis. springer science & business media.

Doucet, A., De Freitas, N., Gordon, N., 2001. An introduction to sequential monte carlo methods, in: Sequential Monte Carlo methods in practice. Springer, pp. 3–14.

Färe, R., 1991. Measuring Farrell efficiency for a firm with intermediate inputs. Academia Economic Papers 19, 329–340.

Färe, R., Grosskopf, S., 1996. Productivity and intermediate products: A frontier approach. Economics letters 50, 65–70.

Fernandez, C., Koop, G., Steel, M.F., 2005. Alternative efficiency measures for multiple-output production. Journal of Econometrics 126, 411–444.

Fernandez, C., Koop, G., Steel, M.F.J., 2002. Multiple-output production with undesirable outputs: an application to nitrogen surplus in agriculture. Journal of the American Statistical Association 97, 432–442.

Flury, T., Shephard, N., 2011. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. Econometric Theory 27, 933–956.

Gallant, A.R., Golub, G.H., 1984. Imposing curvature restrictions on flexible functional forms. Journal of Econometrics 26, 295–321.

Geman, S., Geman, D., 1987. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, in: Readings in computer vision. Elsevier, pp. 564–584.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. Bayesian statistics 4, 641–649.

Girolami, M., Calderhead, B., 2011. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 123–214.

Gordon, N., 1997. A hybrid bootstrap filter for target tracking in clutter. IEEE Transactions on Aerospace and Electronic Systems 33, 353–358.

Gordon, N.J., Salmond, D.J., Smith, A.F., 1993. Novel approach to nonlinear/non-gaussian bayesian state estimation, in: IEE proceedings F (radar and signal processing), IET. pp. 107–113.

Hall, J., Pitt, M.K., Kohn, R., 2014. Bayesian inference for nonlinear structural time series models. Journal of Econometrics 179, 99–111.

Johnes, G., 2013. Efficiency in english higher education institutions revisited: A network approach. Economics Bulletin 33, 2698–2706.

Johnes, G., Soo, K.T., 2017. Grades across universities over time. The Manchester School 85, 106–131.

Johnes, J., 2014. Efficiency and mergers in english higher education 1996/97 to 2008/9: Parametric and non-parametric estimation of the multi-input multi-output distance function. The Manchester School 82, 465–487.

Kotz, S., Balakrishnan, N., Johnson, N., 2000. Continuous multivariate distributions volume 1, second version.

Lee, B.L., Johnes, J., 2022. Using network dea to inform policy: The case of the teaching quality of higher education in England. Higher Education Quarterly 76, 399–421.

Lin, M.T., Zhang, J.L., Cheng, Q., Chen, R., 2005. Independent particle filters. Journal of the American Statistical Association 100, 1412–1421.

Liu, J., West, M., 2001. Combined parameter and state estimation in simulation-based filtering, in: Sequential Monte Carlo methods in practice. Springer, pp. 197–223.

Lozano, S., Adenso-Diaz, B., 2018. Network dea-based biobjective optimization of product flows in a supply chain. Annals of Operations Research 264, 307–323.

Mason, G., Williams, G., Cranmer, S., 2009. Employability skills initiatives in higher education: what effects do they have on graduate labour market outcomes? Education Economics 17, 1–30.

Moreno, P., Lozano, S., 2014. A network dea assessment of team efficiency in the nba. Annals of Operations Research 214, 99–124.

Osiewalski, J., Steel, M.F., 1998. Numerical tools for the Bayesian analysis of stochastic frontier models. Journal of Productivity Analysis 10, 103–117.

Pitt, M.K., dos Santos Silva, R., Giordani, P., Kohn, R., 2012. On some properties of markov chain monte carlo simulation methods based on the particle filter. Journal of Econometrics 171, 134–151.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. Journal of the American statistical association 94, 590–599.

Ramabhadran, V., 1951. A multivariate gamma-type distribution. Sankhyā: The Indian Journal of Statistics , 45–46.

Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the kalman filter. IEEE Aerospace and Electronic Systems Magazine 19, 37–38.

Rodríguez, S.V., Plà, L.M., Faulin, J., 2014. New opportunities in operations research to improve pork supply chain efficiency. Annals of Operations Research 219, 5–23.

Smith, J., McKnight, A., Naylor, R., 2000. Graduate employability: policy and performance in higher education in the UK. The Economic Journal 110, 382–411.

Suleman, F., 2018. The employability skills of higher education graduates: insights into conceptual frameworks and methodological options. Higher Education , 1–16.

Thanassoulis, E., Kortelainen, M., Johnes, G., Johnes, J., 2011. Costs and efficiency of higher education institutions in england: a dea analysis. Journal of the operational research society 62, 1282–1297.

Tierney, L., 1994. Markov chains for exploring posterior distributions. the Annals of Statistics , 1701–1728.

Tone, K., Tsutsui, M., 2009. Network DEA: A slacks-based measure approach. European journal of operational research 197, 243–252.

Zellner, A., 1971a. Bayesian and non-bayesian analysis of the log-normal distribution and log-normal regression. Journal of the American Statistical Association 66, 327–330.

Zellner, A., 1971b. An introduction to Bayesian inference in econometrics. volume 156. Wiley New York.

Appendix A: Markov Chain Monte Carlo

We start with the posterior distribution whose density is:

$$p(\theta,\Sigma,\Omega,\boldsymbol{\mu},\{\sigma_m^2\}_{m=1}^M|\mathcal{Y}) \propto |\Sigma|^{-(nT+\underline{\nu}+1)/2}\cdot\int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}}\exp\left\{-\tfrac{1}{2}\text{tr}\left[\underline{A}+\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})\right]\Sigma^{-1}\right\}\cdot$$
$$C(\Omega)\cdot|\Omega|^{-(nT+\underline{\nu}+1)/2}\exp\left\{-\tfrac{1}{2}\text{tr}\left[\underline{A}+\mathbf{B}(\theta,\mathbf{u})\right]\Omega^{-1}\right\}\sigma_{\mu 1}^{-(nT+\underline{\nu}+1)}\prod_{j=1}^3\sigma_{\mu j}^{-n}\exp\left\{-\frac{\underline{q}+\sum_{i=1}^n\mu_{ij}^2}{2\sigma_j^2}\right\}d\mathbf{u}d\boldsymbol{\mu}. \tag{A.1}$$

From (5)-(7) we can write each translog equation as follows:

$$\begin{aligned} Y_{it,1} &= \mu_{i1}+\mathbf{x}_{it,1}'\theta_1+v_{it,1}-u_{it,1},\\ &\ldots\\ Y_{it,M} &= \mu_{iM}+\mathbf{x}_{it,M}'\theta_M+v_{it,M}-u_{it,M}, \end{aligned} \tag{A.2}$$

assuming the general case of M equations. Stacking all-time observations we obtain:

$$\begin{aligned} \mathbf{Y}_{i,1} &= \mu_{i1}\iota_T+\mathbf{X}_{i1}\theta_1+\mathbf{v}_{i1}-\mathbf{u}_{i1},\\ &\ldots\\ \mathbf{Y}_{i,M} &= \mu_{iM}\iota_T+\mathbf{X}_{iM}\theta_M+\mathbf{v}_{iM}-\mathbf{u}_{iM}, \end{aligned} \tag{A.3}$$

where $\mathbf{Y}_{i,m}=[Y_{i1,m},Y_{i2,m},...,Y_{iT,m}]'$ $(m=1,...,M)$, ι_T is a $T\times 1$ vector of ones, $\mathbf{X}_{im}$ is a $T\times k_m$ matrix of all time observations for $\mathbf{x}_{it,m}$ $(m=1,...,M\ )$, $\mathbf{v}_{im}=[v_{it,m},t=1,...,T]'$, $\mathbf{u}_{im}=[u_{it,m},t=1,...,T]'$ and parameters θ_m are $k_m\times 1$ $(m=1,...,M)$. Collecting observations for all cross sections we have:

$$\begin{aligned} \mathbf{Y}_1 &= \boldsymbol{\mu}_1\otimes\iota_T+\mathbf{X}_1\theta_1+\mathbf{v}_1-\mathbf{u}_1,\\ &\ldots\\ \mathbf{Y}_M &= \boldsymbol{\mu}_M\otimes\iota_T+\mathbf{X}_M\theta_M+\mathbf{v}_M-\mathbf{u}_M, \end{aligned} \tag{A.4}$$

where $\mathbf{Y}_m=[\mathbf{Y}_{1,m}',...,\mathbf{Y}_{n,m}']'$ $(m=1,...,M)$, $\mathbf{X}_m=[\mathbf{X}_{1m}',...,\mathbf{X}_{nm}']'$, $\mathbf{v}_m=[\mathbf{v}_{im}',i=1,...,n]'$ $(m=1,...,M)$, $\mathbf{u}_m=[\mathbf{u}_{im}',i=1,...,n]'$ $(m=1,...,M)$, and $\mu_m=[\mu_{im},i=1,...,n]'$ $(m=1,...,M)$. Therefore, we obtain:

$$\mathbf{Y}=\boldsymbol{\mu}\otimes\iota_T+\mathbf{X}\theta+\mathbf{v}-\mathbf{u}, \tag{A.5}$$

where $\mathbf{Y}=[\mathbf{Y}_1',...,\mathbf{Y}_M']'$, $\boldsymbol{\mu}=[\boldsymbol{\mu}_1',...,\boldsymbol{\mu}_M']'$, $\mathbf{X}=\text{diag}[\mathbf{X}_1,...,\mathbf{X}_M]$, $\vartheta=[\theta_1',...,\theta_M']'$ and $\mathbf{v},\mathbf{u}$ are defined in the obvious way.

A draw from the conditional posterior distribution of θ can be realized as:

$$\theta\sim\mathcal{N}(\hat{\theta},V),\ \theta\in\mathcal{R}, \tag{A.6}$$

where $\hat{\theta}=[\mathbf{X}'(\Sigma^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}[\mathbf{X}'(\Sigma^{-1}\otimes\mathbf{I})\boldsymbol{\psi}]$, $V=[\mathbf{X}'(\Sigma^{-1}\otimes\mathbf{I})\mathbf{X}]^{-1}$, and $\boldsymbol{\psi}=\mathbf{Y}-\boldsymbol{\mu}\otimes\iota_T+\mathbf{u}$. The truncation to the regularity region $\mathcal{R}$ (see (17)) is accomplished by rejection for each MCMC

drawn to maintain the correct posterior.

The conditional posterior distribution of Σ is given by:

$$p(\Sigma|\cdot) \propto |\Sigma|^{-nT+\underline{\nu}+1)/2} \exp\left\{-\tfrac{1}{2}\text{tr}(\underline{A} + \mathbf{A}(\theta, \boldsymbol{\mu}, \mathbf{u}))\Sigma^{-1}\right\}. \tag{A.7}$$

This is a Wishart distribution from which random number generation is straightforward (Zellner, 1971, p. 389).

The conditional posterior distribution of Ω is given by:

$$p(\Omega|\cdot) \propto C(\Omega)^{nT} \cdot |\Omega|^{-nT+\underline{\nu}+1)/2} \exp\left\{-\tfrac{1}{2}\text{tr}(\underline{A} + \mathbf{B}(\theta, \mathbf{u}))\Omega^{-1}\right\}. \tag{A.8}$$

The distribution is not in any known family. However, we can draw a candidate Ω^c from the Wishart distribution whose kernel is:

$$f(\Omega) \propto |\Omega|^{-nT+\underline{\nu}+1)/2} \exp\left\{-\tfrac{1}{2}\text{tr}(\underline{A} + \mathbf{B}(\theta, \mathbf{u}))\Omega^{-1}\right\}. \tag{A.9}$$

Given the existing draw, say $\Omega^{(s)}$ we accept the candidate with probability:

$$\min\left\{1, \left[\frac{C(\Omega^c)}{C(\Omega^{(s)})}\right]^{nT}\right\}, \tag{A.10}$$

Following Tierney (1994), with the probability above we set $\Omega^{(s+1)} = \Omega^c$ else we stay at the previous draw: $\Omega^{(s+1)} = \Omega^{(s)}$. The acceptance rate of this procedure was quite satisfactory - with a median of 92% and never below 80% - in our Monte Carlo experiments and the empirical application. The constant of integration, $C(\Omega)$ is computed by simulation for each Ω using 10,000 randomly generated samples from $\mathbf{z} \sim \mathcal{N}_k(0, \Omega)$ subject to the constraints $P^{-1}\mathbf{z} \geq \mathbf{0}_k$.

To generate the $\boldsymbol{\mu}_i$s $(i = 1, ..., n)$ we can complete the square in the posterior or use Theil's mixed estimator in the following system:

$$\begin{gathered}\mathbf{W}_{i,m} \equiv \mathbf{Y}_{i,m} - \mathbf{X}_{im}\theta_m + \mathbf{u}_{im} = \mu_{im} + \mathbf{v}_m, m = 1, ..., M \\ \mathbf{0}_M = \boldsymbol{\mu}_i + \xi_i,\ \xi_i \sim \mathcal{N}(0, \Phi),\end{gathered} \tag{A.11}$$

where $\Phi = \text{diag}[\sigma^2_{\mu 1}, ..., \sigma^2_{\mu M}]$. Defining $\mathbf{W}_i = [\mathbf{W}_{i,1}, ..., \mathbf{W}_{i,M}]'$ it can be shown that the $M \times 1$ vector $\boldsymbol{\mu}_i$ has the following conditional posterior:

$$\boldsymbol{\mu}_i \sim \mathcal{N}_M(\hat{\boldsymbol{\mu}}_i, \hat{V}),\ i = 1, ..., n, \tag{A.12}$$

where $\hat{\boldsymbol{\mu}}_i = \left[\mathbf{D}'\left(\Sigma^{-1} \otimes \mathbf{I}\right)\mathbf{D} + \Phi^{-1}\right]^{-1}\mathbf{D}'\left(\Sigma^{-1} \otimes \mathbf{I}\right)\mathbf{W}_i$, and $\hat{V} = \left[\mathbf{D}'\left(\Sigma^{-1} \otimes \mathbf{I}\right)\mathbf{D} + \Phi^{-1}\right]^{-1}$. Here, $\mathbf{D}$ is a $T \times (M+1)$ matrix, $\mathbf{D} = I_{M+1} \otimes \iota_T$.

The scale constants have standard conditional posterior distributions:

$$\frac{\underline{q} + \sum_{i=1}^{n} \mu_{im}^2}{\sigma_{\mu m}^2} \sim \chi^2(n + \underline{\nu}),\ m = 1, ..., M. \tag{A.13}$$

It remains to draw from the conditional posteriors of latent inefficiencies. Drawing from the conditional distribution of $\mathbf{u}_{it}$ can be performed by completing the square in (14) to yield:

$$\mathbf{u}_{it}|\theta, \Sigma, \Omega, \{u_{j\tau}, j \neq i, \tau \neq t\} \sim \mathcal{N}_{+}(\mathbf{m}_{it}, \mathbf{V}), \tag{A.14}$$

where $\mathbf{m}_{it} = -(\Sigma^{-1} + \Omega^{-1})^{-1}\Sigma^{-1}\mathbf{v}_{it}(\theta)$, and $\mathbf{V} = (\Sigma^{-1} + \Omega^{-1})^{-1}$. Random drawings can be computed using a Gibbs sampler to draw $u_{it,1}|u_{it,2}, u_{ut,3}, \theta, \Sigma, \Omega$, followed by $u_{it,2}|u_{it,1}, u_{ut,3}, \theta, \Sigma, \Omega$, and $u_{it,3}|u_{it,1}, u_{it,2}, \theta, \Sigma, \Omega$.

In addition to this MCMC procedure we use a more efficient Riemannian MCMC procedure based on Girolami and Calderhead (2011). The Riemannian MCMC procedure uses first and second derivative information from the log posterior and was tested in artificial data sets and the empirical application. It was found to perform better in terms of autocorrelation and convergence and, therefore, we report results based on this procedure. Differences of results between the two procedures were, however, trivial.

Appendix B: Maximum Simulated Likelihood (MSL)

We repeat here the expression for the likelihood function in (15):

$$L(\theta,\Sigma,\Omega,\{\sigma_m^2\}_{m=1}^M;\mathcal{Y}) \propto |\Sigma|^{-nT/2}|\Omega|^{-nT/2}\int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}} \exp\left\{-\tfrac{1}{2}\text{tr}\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})\Sigma^{-1}\right\}\cdot \\ C(\Omega)\cdot\exp\left\{-\tfrac{1}{2}\text{tr}\mathbf{B}(\theta,\mathbf{u})\Omega^{-1}\right\}\sigma_{\mu 1}^{-nT}\prod_{j=1}^3\sigma_{\mu j}^{-n}\exp\left\{-\frac{\sum_{i=1}^n\mu_{ij}^2}{2\sigma_j^2}\right\}d\mathbf{u}d\boldsymbol{\mu}. \tag{B.1}$$

It is possible to concentrate with respect to the different elements and obtain:

$$L(\theta,\Sigma,\Omega,\{\sigma_m^2\}_{m=1}^M;\mathcal{Y}) \propto |\Omega|^{-nT/2}\int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}} |\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})|^{-nT/2}\cdot \\ C(\Omega)\cdot\exp\left\{-\tfrac{1}{2}\text{tr}\mathbf{B}(\theta,\mathbf{u})\Omega^{-1}\right\}\sigma_{\mu 1}^{-nT}\prod_{j=1}^3\sigma_{\mu j}^{-n}\exp\left\{-\frac{\sum_{i=1}^n\mu_{ij}^2}{2\sigma_j^2}\right\}d\mathbf{u}d\boldsymbol{\mu}. \tag{B.2}$$

For economy in notation, we define $\vartheta = \left[\theta', \text{vech}(\Sigma)', \text{vech}(\Omega)', \{\sigma_m^2\}_{m=1}^M\right]'$ and we express the likelihood as follows:

$$L(\vartheta;\mathcal{Y}) \propto \int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}} g(\vartheta,\mathbf{u},\boldsymbol{\mu})d\mathbf{u}d\boldsymbol{\mu}, \tag{B.3}$$

where

$$g(\vartheta,\mathbf{u},\boldsymbol{\mu}) = |\Omega|^{-nT/2}|\mathbf{A}(\theta,\boldsymbol{\mu},\mathbf{u})|^{-nT/2}\cdot \\ C(\Omega)\cdot\exp\left\{-\tfrac{1}{2}\text{tr}\mathbf{B}(\theta,\mathbf{u})\Omega^{-1}\right\}\sigma_{\mu 1}^{-nT}\prod_{j=1}^3\sigma_{\mu j}^{-n}\exp\left\{-\frac{\sum_{i=1}^n\mu_{ij}^2}{2\sigma_j^2}\right\}. \tag{B.4}$$

The major computational burden is in evaluating the multivariate integral with respect to $\mathbf{u}$ and $\boldsymbol{\mu}$. Suppose we had an importance density function $\mathcal{I}_1(\mathbf{u};\gamma)\cdot\mathcal{I}_2(\boldsymbol{\mu};\gamma)$ from which simulation of random vectors $\left\{\mathbf{u}_t^{(s)},\boldsymbol{\mu}_t^{(s)}, s=1,...,S\right\}$ is possible and γ is a known vector of parameters set in advance. In this case we can approximate the likelihood as follows:

$$L(\vartheta;\mathcal{Y}) \propto \int_{\mathbb{R}^{nT}\times\mathbb{R}^{Mn}} \frac{g(\vartheta,\mathbf{u},\boldsymbol{\mu})}{\mathcal{I}_1(\mathbf{u};\gamma)\cdot\mathcal{I}_2(\boldsymbol{\mu};\gamma)}\mathcal{I}_1(\mathbf{u};\gamma)\cdot\mathcal{I}_2(\boldsymbol{\mu};\gamma)d\mathbf{u}d\boldsymbol{\mu}. \tag{B.5}$$

In turn the likelihood can be approximated by simulation:

$$\hat{L}(\vartheta;\mathcal{Y}) \propto S^{-1}\sum_{s=1}^S \frac{g(\vartheta,\mathbf{u}^{(s)},\boldsymbol{\mu}^{(s)})}{\mathcal{I}_1(\mathbf{u}^{(s)};\gamma)\cdot\mathcal{I}_2(\boldsymbol{\mu}^{(s)};\gamma)}. \tag{B.6}$$

In turn, we can maximize the log of the likelihood, $\log\hat{L}(\vartheta;\mathcal{Y})$, with respect to ϑ. To craft the importance densities we rely on information from MCMC. Specifically, we specify first $\mathcal{I}_1(\mathbf{u};\gamma)$, $\mathcal{I}_2(\boldsymbol{\mu};\gamma)$ as

$$\mathcal{I}_1(\mathbf{u};\gamma) = \prod_{i=1}^n\prod_{t=1}^T f_{u,it}(u_{it};\gamma),\ \mathcal{I}_2(\boldsymbol{\mu};\gamma) = \prod_{i=1}^n\prod_{m=1}^M f_{\mu,im}(\mu_{im};\gamma). \tag{B.7}$$

For $f_{u,it}(u_{it};\gamma)$ we use the MCMC draws $\left\{\log u_{it}^{(s)}, s=1,...,S\right\}$ and fit a Student-t distribu-

	Multivariate Half-Normal				Multivariate Exponential			
	stage A	stage B	stage C	overall	stage A	stage B	stage C	overall
post. mean	0.928	0.898	0.950	0.917	0.927	0.933	0.981	0.925
post. median	0.936	0.912	0.955	0.919	0.940	0.948	0.987	0.923
post s.d.	0.028	0.044	0.023	0.071	0.030	0.054	0.023	0.090
5%	0.839	0.745	0.882	0.780	0.839	0.788	0.923	0.798
95%	0.977	0.976	0.985	0.973	0.970	0.989	0.995	1.000

Table B.1: Sample statistics of posterior mean efficiencies

tion with five degrees of freedom with location and scale parameters determined from the draws $\left\{u_{it}^{(s)}, s = 1, ..., S\right\}$. For $f_{\mu,im}(\mu_{im}; \gamma)$ we use a Student-t distribution with five degrees of freedom with location and scale parameters determined from the draws $\left\{\mu_{im}^{(s)}, s = 1, ..., S\right\}$. The degrees of freedom and location and scale parameters of these importance densities are subsumed in parameter vector γ. Without information from MCMC we have found it difficult to find appropriate importance densities as parameters γ must be fixed and are not allowed to change in thew course of MSL. This testifies to the difficulty of the problem of MSL and the fact that MCMC methods of inference are better suited. We try S =100, 200, 300 etc until estimates of ϑ converge within 10^{-6}. Finally, we select $S = 700$ in the case of the multivariate half - normal and S=500 in the case of the multivariate exponential distribution. The empirical results are reported in Tables B.1 and B.2.

Appendix C: Description of Sequential Monte Carlo

The particle filter methodology can be applied to state space models of the general form:

$$y_T \sim p(y_t|x_t), \;\; s_t \sim p(s_t|s_{t-1}), \tag{C.1}$$

where s_t is a state variable ($\mathbf{w}_{it}$ in our case). For general introductions see Gordon (1997), Gordon et al. (1993), Doucet et al. (2001) and Ristic et al. (2004).

Given the data Y_t the posterior distribution $p(s_t|Y_t)$ can be approximated by a set of (auxiliary) particles $\left\{s_t^{(i)}, i = 1, ..., .N\right\}$ with probability weights $\left\{w_t^{(i)}, i = 1, ..., N\right\}$ where $\sum_{i=1}^{N} w_t^{(i)} = 1$. The predictive density can be approximated by:

$$p(s_{t+1}|Y_t) = \int p(s_{t+1}|s_t)p(s_t|Y_t)ds_t \simeq \sum_{i=1}^{N} p(s_{t+1}|s_t^{(i)})w_t^{(i)}, \tag{C.2}$$

	Multivariate Half-Normal			Multivariate Exponential		
	stage A	stage B	stage C	stage A	stage B	stage C
intake quality	0.475	0.387	0.684	0.347	0.360	0.655
	(0.044)	(0.071)	(0.244)	(0.042)	(0.055)	(0.223)
staff:student ratio	0.375	0.267	0.381	0.377	0.265	0.370
	(0.043)	(0.055)	(0.032)	(0.039)	(0.048)	(0.032)
per student spend	0.251	0.258	0.004	0.226	0.263	0.003
	(0.021)	(0.020)	(0.017)	(0.016)	(0.019)	(0.022)
x_2, research reputation	-	0.177	-	-	0.188	-
		(0.062)			(0.068)	
y_1, degree results	-	0.157	-	-	0.158	-
		(0.078)			(0.068)	
σ_v	0.057	0.051	0.035			
	(0.008)	(0.013)	(0.006)			
σ_u	0.080	0.123	0.060			
	(0.017)	(0.019)	(0.014)			
α_m				1.319	1.225	1.389
				(0.039)	(0.019)	(0.031)
α_0				0.984 (0.030)		

Notes: Standard deviation in parentheses.

Table B.2: Marginal Effects

and the final approximation for the filtering density is:

$$p(s_{t+1}|Y_t) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|Y_t) \simeq p(y_{t+1}|s_{t+1}) \sum_{i=1}^{N} p(s_{t+1}|s_t^{(i)}) w_t^{(i)}. \tag{C.3}$$

The basic mechanism of particle filtering rests on propagating $\left\{s_t^{(i)}, w_t^{(i)}, i = 1, \ldots, N\right\}$ to the next step, viz. $\left\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i = 1, \ldots, N\right\}$ but this often suffers from the weight degeneracy problem. If parameters $\theta \in \Theta \in \Re^k$ are available, as is often the case, we follow Liu and West (2001) where parameter learning takes place via a mixture of multivariate normals:

$$p(\theta|Y_t) \simeq \sum_{i=1}^{N} w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2 V_t), \tag{C.4}$$

where $\bar{\theta}_t = \sum_{i=1}^{N} w_t^{(i)} \theta_t^{(i)}$, and $V_t = \sum_{i=1}^{N} w_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0, 1)$ as $a = (1 - b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also Casarin and Marin (2007).

Andrieu and Roberts (2009), Flury and Shephard (2011) and Pitt et al. (2012) provide the Particle Metropolis-Hastings (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y|\theta)$ as $p(Y|\theta)$ is often not available in closed form.

Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say $L^j = \hat{p}_N(Y|\theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c|\theta^{(j)})$ yielding $L^c = \hat{p}_N(Y|\theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min\left\{1,\ \frac{p(\theta^c)L^c}{p(\theta^{(j)}L^j}\frac{q(\theta^{(j)}|\theta^c}{q(\theta^c|\theta^{(j)})}\right\}, \tag{C.5}$$

otherwise we repeat the current draws: $\left\{\theta^{(j+1)}, L^{j+1}\right\} = \left\{\theta^{(j)}, L^j\right\}$.

Hall et al. (2014)propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering Pitt et al. (2012) used within PMCMC requires far fewer particles that the standard particle filtering algorithm to approximate $p(Y|\theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t|s_{t-1})$. When this is not possible, Hall et al. (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have:

$$p(y_t|s_{t-1}) = \int p(y_t|s_t)p(s_t|s_{t-1})ds_t, \tag{C.6}$$

$$p(u_t|s_{t-1}; y_t) = p(y_t|s_{t-1}, u_t)p(u_t|s_{t-1})/p(y_t|s_{t-1}). \tag{C.7}$$

If one can evaluate $p(y_t|s_{t-1})$ and simulate from $p(u_t|s_{t-1}; y_t)$ the filter would be fully adaptable Pitt and Shephard (1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t|s_{t-1})$ by matching the first two moments of $p(y_t|s_{t-1})$. So in some way we find that the approximating density $p(y_t|s_{t-1}) = N\left(\mathbb{E}(y_t|s_{t-1}), \mathbb{V}(y_t|s_{t-1})\right)$. In the second stage, we know that $p(u_t|y_t, s_{t-1}) \propto p(y_t|s_{t-1}, u_t)p(u_t)$. For $p(u_t|y_t, s_{t-1})$ we know it is multimodal so suppose it has M modes are $\hat{u}_t^m$, for $m = 1, \ldots, M$. For each mode we can use a Laplace approximation. Let $l(u_t) = log\left[p(y_t|s_{t-1}, u_t)p(u_t)\right]$. From the Laplace approximation we obtain:

$$l(u_t) \simeq l(\hat{u}_t^m) + \tfrac{1}{2}(u_t - \hat{u}_t^m)'\nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m). \tag{C.8}$$

Then we can construct a mixture approximation:

$$g(u_t|x_t, s_{t-1}) = \sum_{m=1}^{M} \lambda_m (2\pi)^{-d/2}|\Sigma_m|^{-1/2} \exp\left\{\tfrac{1}{2}(u_t - \hat{u}_t^m)'\Sigma_m^{-1}(u_t - \hat{u}_t^m\right\}, \tag{C.9}$$

where $\Sigma_m = -[\nabla^2 l(\hat{u}_t^m)]^{-1}$ and $\lambda_m \propto \exp\left\{l(u_t^m)\right\}$ with $\sum_{m=1}^{M} = 1$. This is done for each particle s_t^i. This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a proposal for s_t directly from the measurement density $p(y_t|s_t)$ although Hall et al. (2014) are quite right in pointing out that the state equation can be very informative.

In the standard particle filter of Gordon et al. (1993) particles are simulated through the state

density $p(s_t^i|s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t|s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

For $t = 0, \ldots, T-1$ given samples $s_t^k \sim p(s_t|Y_{1:t})$ with mass π_t^k for $k = 1, ..., N$.

1) For $k = 1, \ldots, N$ compute $\omega_{t|t+1}^k = g(y_{t+1}|s_t^k)\pi_t^k$, $\pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^N \omega_{t|t+1}^i$.

2) For $k = 1, \ldots, N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t}^i(ds_t)$.

3) For $k = 1, \ldots, N$ draw $u_{t+1}^k \sim g(u_{t+1}|\tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; maru_{t+1}^k)$.

4) For $k = 1, \ldots, N$ compute

$$\omega_{t+1}^k = \frac{p(y_{t+1}|s_{t+1}^k)p(u_{t+1}^k)}{g(y_{t+1}|s_t^k)g(u_{t+1}^k|\tilde{s}_t^k, y_{t+1})}, \pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}. \tag{C.10}$$

It should be mentioned that the estimate of likelihood from ADPF is:

$$p(Y_{1:T}) = \prod_{t=1}^{T} \left(\sum_{i=1}^{N} \omega_{t-1|t}^i \right) \left(N^{-1} \sum_{i=1}^{N} \omega_t^i \right). \tag{C.11}$$

We use this approach to integrate out the latent dynamic process from (23) and (22).