THE UNIVERSITY OF MELBOURNE

Department of Economics

Working Paper Series

A Stochastic Frontier Model for Discrete Ordinal Outcomes: A Health Production Function

William Griffiths, Xiaohui Zhang
and Xueyan Zhao

February 2010

Research Paper Number 1092

ISSN: 0819-2642
ISBN: 978 0 7340 4445 7

Department of Economics
The University of Melbourne
Parkville VIC 3010
www.economics.unimelb.edu.au

A Stochastic Frontier Model for Discrete Ordinal Outcomes: A Health Production Function

William Griffiths

Department of Economics, University of Melbourne

Xiaohui Zhang

Department of Econometrics and Business Statistics, Monash University

Xueyan Zhao

Department of Econometrics and Business Statistics and Centre for Health Economics, Monash University

25 February 2010

Abstract

The stochastic frontier model used for continuous dependent variables is extended to accommodate output measured as a discrete ordinal outcome variable. Conditional on the inefficiency error, the assumptions of the ordered probit model are adopted for the log of output. Bayesian estimation utilizing a Gibbs sampler with data augmentation is applied to a convenient re-parameterisation of the model. Using panel data from an Australian longitudinal survey, demographic and socioeconomic characteristics are specified as inputs to health production, whereas production efficiency is made dependent on lifestyle factors. Posterior summary statistics are obtained for selected health status probabilities, efficiencies, and marginal effects.

JEL classification: C11, C21, C23, I12

Keywords: Bayesian estimation, Gibbs sampling, ordered probit, production efficiency

Corresponding Author:

William Griffiths
Department of Economics
University of Melbourne
Vic 3010, Australia

Email: wegrif@unimelb.edu.au

Phone: +613 8344 3622

Fax: +613 8344 6899

1. Introduction

Since its introduction by Aigner et al. (1977) and Meeusen and van den Broeck (1977), the stochastic frontier model has been widely used in the analysis of productivity and firm efficiency, and has been extended in numerous directions. Extensions include alternative assumptions for the distribution of the inefficiency error, the use of panel as well as cross-section data, the specification of time-varying inefficiencies that are related to firm characteristics, the introduction of heteroskedasticity, the use of dual cost and profit frontiers as well as production frontiers, and multiple output models. Applications have used both firm level and country level datasets, and have evaluated the performance of production units for both traditional and service industries. Estimation has been carried out from both the sampling theory and Bayesian standpoints. Surveys of different aspects of the literature can be found in Bauer (1990), Kim and Schmidt (2000), and Greene (2005). Of particular relevance to this study is Bayesian estimation of the stochastic frontier model, introduced by van den Broeck et al. (1994) and surveyed by Koop and Steel (2001). Books with substantial reviews of the literature are Kumbhakar and Lovell (2000) and Coelli et al. (2005).

An assumption common to all past studies is that the dependent variable (logarithm of output or cost) is a continuous random variable that is fully observed. In this paper we extend modelling and estimation of the stochastic frontier model to the case where the dependent variable is latent and is observed only as an ordered categorical variable. Conditional on the inefficiency error, an ordered probit model is used to model data of this form. The context in which our model is specified is that of an individual's health production function.

Health economists have considered both the production of health care and the production of health itself. Stochastic production frontier models, together with data envelopment analysis (DEA, see for example Coelli, et al. 2005) have been used to examine the production of health care and to benchmark hospital performance and efficiency. Some examples are Gerdtham, et al. (1999), Rosko (2001), Brown (2003), Street (2003), and Puig-Junoy and Ortun (2004). Studies on the production of health itself can be further divided into macro studies which use aggregate statistics to investigate the health production of a country or region, or a particular group of people in a country, and micro studies which focus on the health of individuals using data at the individual level. For macro-level studies, the output of health production is often measured by continuous variables which are aggregated health indicators of a country such as mortality rate or disability-adjusted life expectancy. Population health expenditure, and aggregated measures of education, lifestyle and environmental factors, are used as health production inputs. Both parametric stochastic frontier and nonparametric DEA approaches are used in these macro studies, and efficiency measures of health production are estimated and compared across countries. See for example Puig-Junoy (1998), Evan et al. (2000a, 2000b), Thornton (2002), Hollingsworth and Wildman (2003) and Fayissa and Gutema (2005).

Empirical research on population health using individual level data has become increasingly important in recent times for both developed and developing countries. Population ageing, labour shortages, epidemic trends for many chronic diseases, and obesity in the context of changing modern lifestyle have all intensified the urgency for government intervention in population health. The relationship between risk factors created via lifestyle behaviour with health outcomes at the individual level is crucial for informing public-funded health campaigns. Correlation between socioeconomic characteristics and health status is also an important measure, particularly in studies of health inequality. In the human capital theory of economics, health is an important endowment of human capital, like education, and is a product of household production

(Muth 1966; Grossman 1972). The level of health is produced with market goods, such as medical care, and an individual's own time and effort via lifestyle behaviour. Empirical studies using individual level survey data include Desai (1987), Atkinson and Crocker (1992), Akin (1992), Contoyannis and Jones (2004), Jacobs et. al (2004) and Hakinen et al. (2006).

A commonly used measure for an individual's health status is the self-assessed overall health grade measured as ordered multiple discrete choices. Information is collected from survey questions, such as: "Would you say your health in general is excellent, very good, good, fair or poor?" Although the self-assessed grade is subject to measurement errors as are all other self-reported variables, empirical evidence suggests that it is a reliable measure of overall health status. For example, it is shown that it is a good summary of health conditions in various dimensions of physical, mental, social or functional health (Liang 1986; Jylha 1994; Baron-Epel, et al. 2005) and a valid predictor for mortality (Mossey and Shapiro 1982; Benjamins, et al. 2004; van den Brink, et al. 2005). Although linear regression models have been used to study the self-reported health grade (Desai 1987), econometric models specifically designed for ordered discrete dependent variables, such as the ordered probit model, are more suitable for analysing health production functions of this type (Contoyannis and Jones 2004). However, unlike studies of health production using macro data, the issue of production efficiency has not been examined for individual health production, and has not appeared in the production frontier literature. In empirical research of individual health, it is common to find that persons with similar demographic and socioeconomic characteristics report completely different statements about their personal health status, because of different lifestyle behaviour. It is thus a natural step to extend the stochastic production frontier model to a discrete ordered dependent variable so that the techniques in that literature can be used to study the efficiency of individual level health production.

In this paper, we extend the stochastic production frontier model to ordered discrete dependent variables, and apply the model to study the production and production efficiency of individual level health using panel data from an Australian longitudinal survey. We study the effects of socioeconomic and demographic characteristics on health production and parameterise the model to allow the efficiency of health production to vary by lifestyle factors. We use a Bayesian approach with a Markov chain Monte Carlo (MCMC) algorithm for estimation and inference, adopting and modifying previous Bayesian algorithms for stochastic frontier models with continuous outputs (van den Broeck, et al. 1994) and for ordered probit models (Albert and Chib 1993; Nandram and Chen 1996; Li and Tobias 2006).

The structure of this paper is as follows. In Section 2 the econometric framework for modelling a stochastic frontier with discrete ordinal output with cross-sectional data is introduced. The model for panel data is introduced in Section 3. Quantities of interest, including estimated probabilities, marginal effects and efficiency measures are discussed in Section 4. In Section 5 the model for panel data is applied to health production. Section 6 contains a short conclusion.

2. Modelling with cross-sectional data

2.1 Model specification

We first consider a production frontier model for cross-sectional data with sample size *N* where y_i is a discrete observable random variable that takes one of $J+1$ ordered values from 0 to *J*. As in the standard ordered probit model, an unobserved continuous latent variable y_i^* can be mapped to the observed discrete outcome y_i via some boundary parameters. The latent production output variable y_i^* is assumed positive and is related to a $1\times(k+1)$ vector of input variables x_i, whose first element is unity. Following the typical stochastic frontier model setup, we write the natural logarithm of latent variable y_i^* as

$$\ln y_i^* = f(x_i, \beta) + v_i - u_i, \qquad i = 1, 2, \ldots, N. \tag{1}$$

The $(k+1)\times 1$ vector β contains unknown parameters. The v_i's are independent identically distributed symmetric errors that follow a standard normal distribution, i.e., $v_i \sim \text{i.i.d. } N(0,1)$. They reflect measurement and specification errors. The assumption of unit variance is in line with that needed for identification in the traditional ordered probit model. The u_i's are independent identically distributed non-negative error terms, with a given u_i measuring the inefficiency level of firm *i*. The value $u_i = 0$ indicates technical efficiency, while $u_i > 0$ is an indication of technical inefficiency where production of the *i*-th firm lies below the production frontier. Technical efficiency is defined as $r = e^{-u}$, with $0 < r \le 1$. The errors v_i and u_i are assumed independent. Several one-sided distributions of u_i have been considered in the literature. Early work on the stochastic frontier model presented by Meeusen and van den Broeck (1977) adopts an exponential distribution. Aigner et al. (1997) assume u_i follows a half normal distribution. Other suggestions include truncated normal (Stevenson 1980) and gamma (Greene 1990) distributions. In this paper, we assume u_i follows an exponential distribution, i.e., $u_i \sim \text{i.i.d. } \Gamma\left(1, \lambda^{-1}\right)$, with mean λ and variance λ^2.

Assuming that actual unobserved output is positive, and specifying the boundaries or thresholds as $1, \mu_1, \mu_2, \ldots, \mu_{J-1}$, the mapping between y_i and y_i^* can be written as

$$y_i = \begin{cases} 0 & 0 < y_i^* \le 1 \\ 1 & 1 < y_i^* \le \mu_1 \\ & \vdots \\ j & \mu_{j-1} < y_i^* \le \mu_j \qquad j = 2, \ldots, J-1 \\ & \vdots \\ J & y_i^* > \mu_{J-1}. \end{cases} \tag{2}$$

Since the dependent variable in (1) is $\ln y_i^*$ rather than y_i^*, it is convenient to rewrite the mapping in (2) in terms of the logarithms of y_i^* and its thresholds μ_j's. Specifically, let $g_i^* = \ln(y_i^*)$, $\gamma_j = \ln(\mu_j)$, $j = 1, 2, \ldots, J-1$, and set $\gamma_{-1} = -\infty$, $\gamma_0 = 0$ and $\gamma_J = +\infty$. Then (1) can be written as

$$g_i^* = f(x_i, \beta) + v_i - u_i, \tag{3}$$

and (2) becomes

$$y_i = j, \text{ if and only if } \gamma_{j-1} < g_i^* \le \gamma_j, \quad j = 0, 1, \ldots, J \tag{4}$$

The representation in (4) is the specification commonly used for the ordered probit model. The thresholds $\gamma_1, \gamma_2, \ldots, \gamma_{J-1}$ are unknown parameters that need to be estimated along with β. In the special case where $f(x_i, \beta)$ is linear, for example, a Cobb-Douglas production technology where the x_i are the logarithms of the inputs, (3) can be written as

$$g_i^* = x_i\beta + v_i - u_i. \tag{5}$$

It is convenient at this point to introduce some matrix notation. In what follows we use the following definitions: $y = (y_1, y_2, \ldots, y_N)'$, $g^* = (g_1^*, g_2^*, \ldots, g_N^*)'$, $v = (v_1, v_2, \ldots, v_N)'$, $u = (u_1, u_2, \ldots, u_N)'$, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{J-1})'$, and $X = (x_1', x_2', \ldots, x_N')'$ is an $N \times (k+1)$ matrix whose first column contains ones.

Our objective is to describe how Bayesian estimation of this model can be carried out. The first step in this direction is to specify conditional posterior densities which can be used in a Gibbs sampling algorithm to draw values from the joint posterior density of the unknown parameters β, γ and λ, and the latent variables g^* and *u*. Later, we describe how these draws can be used to get posterior densities on other quantities of interest such as probabilities for each level of production conditional on x_i, and various measures of efficiency. In our specification of the conditional posterior densities for β, γ, λ, g^* and *u*, we begin by reviewing two alternative algorithms that have been suggested in the literature for the ordered probit model (where $u = 0$ and only β, γ and g^* are relevant), and then introduce the extra conditional posterior densities necessary to accommodate $u \geq 0$. Conditional posterior densities are also specified for a further extension where λ is allowed to vary over individuals depending on another set of exogenous variables.

2.2 Conditional posterior densities for an ordered probit production model

In this section two MCMC algorithms for estimating the traditional ordered probit model are reviewed prior to introducing the complications caused by the stochastic frontier inefficiency error. Our review borrows much from Chen et al. (2000, Ch.2).

A standard ordered probit production function without the inefficiency error component $(u_i \equiv 0,\ i = 1, \ldots, N)$ can be written as

$$g_i^* = x_i\beta + v_i$$

$$y_i = j, \text{ if and only if } \gamma_{j-1} < g_i^* \leq \gamma_j, \quad j = 0,1,\dots,J \tag{6}$$

$$\gamma_{-1} = -\infty, \ \gamma_0 = 0, \text{ and } \gamma_J = +\infty.$$

The probability of $y_i = j$ $(i = 1,\dots,N;\ j = 0,1,\dots,J)$ is given by

$$p_{ij} = \Pr(y_i = j) = \Phi(\gamma_j - x_i\beta) - \Phi(\gamma_{j-1} - x_i\beta)\,, \tag{7}$$

where $\Phi(\bullet)$ denotes the cumulative distribution function for a standard normal random variable. The likelihood function is

$$\begin{aligned} l\left(\beta,\gamma \mid y, X\right) &= \prod_{i=1}^{N}\left\{\sum_{j=0}^{J} p_{ij} I(y_i = j)\right\} \\ &= \prod_{i=1}^{N}\left\{\sum_{j=0}^{J}\Big[\Big(\Phi(\gamma_j - x_i\beta) - \Phi(\gamma_{j-1} - x_i\beta)\Big) I(y_i = j)\Big]\right\} \\ &= \prod_{i=1}^{N}\left\{\Phi(\gamma_{y_i} - x_i\beta) - \Phi(\gamma_{y_i-1} - x_i\beta)\right\}. \end{aligned} \tag{8}$$

where $I(\cdot)$ is the indicator function which is equal to one when its argument is true and zero otherwise. Assuming that $(\beta,\ \gamma)$ has an improper uniform prior, i.e., $p(\beta,\ \gamma) \propto 1$, the posterior density for $(\beta,\ \gamma)$ is proportional to (8). That is,

$$p\left(\beta,\gamma \mid y, X\right) \propto \prod_{i=1}^{N}\left\{\Phi(\gamma_{y_i} - x_i\beta) - \Phi(\gamma_{y_i-1} - x_i\beta)\right\}. \tag{9}$$

To facilitate estimation via the Gibbs sampler, Albert and Chib (1993) treat the latent variables g_i^* as unknown parameters, in which case the posterior density for $(\beta,\ \gamma,\ g^*)$ becomes

$$p\left(\beta,\gamma,g^* \mid y, X\right) \propto \prod_{i=1}^{N}\left[\exp\left(-\frac{1}{2}\left(g_i^* - x_i\beta\right)^2\right) I\left(\gamma_{y_i-1} < g_i^* \leq \gamma_{y_i}\right)\right]. \tag{10}$$

The original Gibbs sampler for the ordered probit model proposed by Albert and Chib (1993) uses the following conditional posterior densities.

The g_i^* $(i=1,\dots,N)$ follow conditionally independent truncated normal distributions

$$\left(g_i^* \mid \gamma, \beta, X, y\right) \sim N\left(x_i\beta,\, 1\right) I\left(\gamma_{y_i-1} < g_i^* \leq \gamma_{y_i}\right). \tag{11}$$

The conditional posterior density for β is the normal distribution

$$\left(\beta \mid \gamma, g^*, X, y\right) \sim N\left[(X'X)^{-1}X'g^*,\ (X'X)^{-1}\right]. \tag{12}$$

The conditional posterior density for γ_j, $(j=1,2,\dots,J-1)$ is the uniform distribution

$$\left(\gamma_j \mid \gamma^{(-j)}, g^*, \beta, X, y\right) \sim U\left(a_j, b_j\right), \tag{13}$$

where $a_j = \max\left\{\gamma_{j-1},\, \max_i(g_i^* \mid y_i = j)\right\}$, $b_j = \min\left\{\gamma_{j+1},\, \min_i(g_i^* \mid y_i = j+1)\right\}$, and $\gamma^{(-j)}$ denotes γ without γ_j.

The Albert-Chib algorithm is a convenient one because drawing from truncated normal, normal and uniform distributions is straightforward. However, when N is large, say greater than 50, convergence of the Gibbs sampler can be slow (Cowles 1996). The interval (a_j, b_j) can be very narrow, leading to values of the threshold parameters γ_j that change very little and are highly correlated in successive iterations. The slow convergence of the γ_j can feed through into slow convergence of β. To overcome this problem alternative algorithms have been suggested by Cowles (1996), Nadram and Chen (1996), Chen and Dey (1996), and Albert and Chib (1998). The Nadram-Chen algorithm, which we utilize for our stochastic frontier model, has two main innovations: (i) a reparameterization of the model eliminates one of the thresholds parameters and introduces a latent-variable variance parameter that is no longer equal to one, and (ii) a Metropolis-Hastings step, with a Dirichlet proposal density, is used to sample from the conditional posterior density for all remaining thresholds, conditional on β and the new variance parameter, but not conditional on the latent variables.

To describe the Nadram-Chen algorithm, we first divide equation (6) by γ_{J-1}:

$$\frac{g_i^*}{\gamma_{J-1}} = x_i \frac{\beta}{\gamma_{J-1}} + \frac{v_i}{\gamma_{J-1}}, \qquad \frac{v_i}{\gamma_{J-1}} \sim N\left(0, \frac{1}{\gamma_{J-1}^2}\right),$$

$$y_i = j, \text{ if and only if } \frac{\gamma_{j-1}}{\gamma_{J-1}} < \frac{g_i^*}{\gamma_{J-1}} \le \frac{\gamma_j}{\gamma_{J-1}}, \quad j = 0,1,\dots,J \tag{14}$$

Then, after defining $\tilde{v}_i = v_i \big/ \gamma_{J-1}$ and setting $\sigma_v^2 = \text{var}(\tilde{v}_i) = 1\big/\gamma_{J-1}^2$, equation (14) suggests the following reparameterization

$$\sigma_v = 1\big/\gamma_{J-1}, \quad \tilde{g}^* = g^*\big/\gamma_{J-1}, \quad \tilde{\beta} = \beta\big/\gamma_{J-1}, \quad \text{and} \quad \tilde{\gamma} = \gamma\big/\gamma_{J-1} \tag{15}$$

From the above transformations, the reparameterized model is $\tilde{g}_i^* = x_i\tilde{\beta} + \tilde{v}_i$ with new thresholds $\tilde{\gamma}_{-1} = -\infty < \tilde{\gamma}_0 = 0 < \tilde{\gamma}_1 < ... < \tilde{\gamma}_{J-2} < \tilde{\gamma}_{J-1} = 1 < \tilde{\gamma}_J = +\infty$. The number of unknown thresholds has been reduced from $J-1$ to $J-2$, and the scale parameter σ_v has been added. When $J = 2$, implying three categories, there are no unknown threshold parameters.

At this point it is convenient to change the prior density, anticipating what will be needed for the MCMC algorithm after the inefficiency term has been introduced. We assume that the prior $p\left(\tilde{\beta}, \tilde{\gamma}\right)$ is uniform, and that $p\left(\sigma_v^2\right)$ is an inverted gamma density with shape parameter a_v and scale parameter b_v. That is, $\sigma_v^2 \sim IG(a_v, b_v)$. This prior is computationally convenient and, if desired, can be made noninformative by suitable choices of a_v and b_v. It leads to the following joint posterior density for the reparameterized model

$$p\left(\tilde{\beta},\tilde{\gamma},\tilde{g}^*,\sigma_v^2 \mid y,X\right) \propto \frac{1}{\left(\sigma_v^2\right)^{a_v+1}} \exp\left(-\frac{b_v}{\sigma_v^2}\right)$$
$$\times \prod_{i=1}^{N}\left[\frac{1}{\sigma_v}\exp\left(-\frac{1}{2\sigma_v^2}\left(\tilde{g}_i^* - x_i\tilde{\beta}\right)^2\right) I\left(\tilde{\gamma}_{y_i-1} < \tilde{g}_i^* \le \tilde{\gamma}_{y_i}\right)\right].$$

The conditional posterior densities for implementing the Gibbs sampler are as follows.

The conditional posterior densities for the $\tilde{g}_i^*$ are independent truncated normal distributions

$$\left(\tilde{g}_i^* \mid \tilde{\gamma},\tilde{\beta},\sigma_v^2,X,y\right) \sim N\left(x_i\tilde{\beta},\sigma_v^2\right) \ I\left(\tilde{\gamma}_{y_i-1} < \tilde{g}_i^* \le \ \tilde{\gamma}_{y_i}\right), \qquad i=1,\ldots,N. \tag{16}$$

The conditional posterior density for $\tilde{\beta}$ is the normal distribution

$$\left(\tilde{\beta} \mid \tilde{g}^*,\sigma_v^2,X,y\right) \sim N\left[(X'X)^{-1}X'\tilde{g}^*,\ \sigma_v^2(X'X)^{-1}\right]. \tag{17}$$

The conditional posterior density for σ_v^2 is the inverted gamma distribution

$$\left(\sigma_v^2 \mid \tilde{g}^*,\tilde{\beta},X,y\right) \sim IG\left(\frac{N}{2}+a_v,\ b_v+\frac{1}{2}\left[\left(\tilde{g}^*-X\tilde{\beta}\right)'\left(\tilde{g}^*-X\tilde{\beta}\right)\right]\right). \tag{18}$$

For $J=2$, there are no unknown thresholds in $\tilde{\gamma}$, and sampling from the conditional densities in (16)-(18) is sufficient. For $J \ge 3$, an extra step is required to sample from the conditional posterior density for the unknown elements in $\tilde{\gamma}$, namely,

$$p\left(\tilde{\gamma} \mid \tilde{\beta},\sigma_v^2,X,y\right) \propto \prod_{i=1}^{N}\left[\Phi\left(\frac{\tilde{\gamma}_{y_i}-x_i\tilde{\beta}}{\sigma_v}\right)-\Phi\left(\frac{\tilde{\gamma}_{y_i-1}-x_i\tilde{\beta}}{\sigma_v}\right)\right]. \tag{19}$$

Because (19) is not a density function whose form is recognizable, a Metropolis-Hastings step is used to draw from it. A Dirichlet proposal density is constructed for $p\left(\tilde{\gamma} \mid \tilde{\beta}, \sigma_v^2, X, y\right)$ in the following way. Let the differences between adjacent thresholds be defined as $q_j = \tilde{\gamma}_j - \tilde{\gamma}_{j-1}$ for $j = 1, 2, \ldots, J-1$, and let $q = (q_1, q_2, \ldots, q_{J-1})'$. Then, $q_j \geq 0$, and $\sum_{j=1}^{J-1} q_j = 1$, making the Dirichlet distribution a possible proposal density for q. Its density is given by

$$p\left(q \mid \tilde{\beta}, \sigma_v^2, X, y\right) \propto \prod_{j=1}^{J-1} q_j^{\alpha_j n_j - 1} \tag{20}$$

where $0 \leq \alpha_j \leq 1,\ j = 1, 2, \ldots, J-1$, are tuning parameters, and $n_j = \sum_{i=1}^{N} I(y_i = j)$ is the number of sample observations in category j. The advantages of the proposal density (20) are that the entire vector q can be drawn at once, and it does not depend on $\tilde{\beta}$ and σ_v^2. The tuning parameters, $\alpha_j \ \ (j = 1, 2, \ldots, J-1)$ are chosen to make the dispersion of the distribution of q comparable to or at least as large as that of the posterior distribution of $\tilde{\gamma}$. To perform the Metropolis-Hastings step a set of candidate values q^{can} is drawn from $p\left(q \mid \tilde{\beta}, \sigma_v^2, X, y\right)$, and transformed to a set of candidate values $\tilde{\gamma}^{can}$. Given values $\tilde{\gamma}$ from the previous iteration, the vector $\tilde{\gamma}^{can}$ is accepted with probability $a = \min\{R, 1\}$ where

$$R = \left\{ \prod_{i=1}^{N} \frac{\Phi\left(\left(\tilde{\gamma}_{y_i}^{can} - x_i\tilde{\beta}\right)/\sigma_v\right) - \Phi\left(\left(\tilde{\gamma}_{y_i-1}^{can} - x_i\tilde{\beta}\right)/\sigma_v\right)}{\Phi\left(\left(\tilde{\gamma}_{y_i} - x_i\tilde{\beta}\right)/\sigma_v\right) - \Phi\left(\left(\tilde{\gamma}_{y_i-1} - x_i\tilde{\beta}\right)/\sigma_v\right)} \right\} \left\{ \prod_{j=1}^{J-1} \left(\frac{q_j}{q_j^{can}} \right)^{\alpha_j n_j - 1} \right\} \tag{21}$$

Values $\tilde{\gamma}$ from the previous iteration are accepted with probability $1-a$. Observations from the posterior densities of the original parameters in (6) are recovered from the rescaled parameters by dividing by σ_v.

2.3 Conditional posteriors for the ordered probit stochastic production frontier

In this section, the restriction $u_i \equiv 0$ $(i = 1,...,N)$ is relaxed to yield the ordered probit frontier model as given in equation (5). A Gibbs sampling algorithm that combines the Nadram-Chen algorithm for the ordered probit model and an algorithm for the stochastic frontier model with continuous output (Koop, et al. 1997; Koop and Steel 2001) is presented. In line with the previous section, we work with a reparameterized model

$$\tilde{g}_i^* = x_i\tilde{\beta} + \tilde{v}_i - \tilde{u}_i \tag{22}$$

where, in addition to the transformations defined in and around equation (15), we have $\tilde{u}_i = u_i/\gamma_{J-1}$ and $E(\tilde{u}_i) = \tilde{\lambda} = E(u_i)/\gamma_{J-1} = \lambda/\gamma_{J-1}$, with $\tilde{u}_i \sim$ i.i.d. $\Gamma\left(1,\tilde{\lambda}^{-1}\right)$. Both $\tilde{g}^*$ and the inefficiency error $\tilde{u}$ are treated as unknown parameters with values being drawn from their conditional posterior densities. Then, conditional on $\tilde{g}^*$ and $\tilde{u}$, the stochastic frontier model for ordinal outcomes in (22) reduces to the standard linear regression model, facilitating draws from the conditional posterior densities for the parameters.

For a prior density we continue to assume all parameters are a prior independent, that $p\left(\tilde{\beta},\tilde{\gamma}\right) \propto 1$, and that $\sigma_v^2 \sim IG(a_v, b_v)$. In addition, we assume that $\tilde{\lambda} \sim IG(a_\lambda, b_\lambda)$. Following van den Broeck, et al. (1994) and many subsequent authors (see, for example, Koop and Steel 2001), the hyperparameters a_λ and b_λ can be set by considering the implied distribution of efficiency $r = \exp\left(-\tilde{u}\right)$. We follow previous tradition and set

$a_\lambda = 1$ and $b_\lambda = -\ln(r^*)$, yielding a prior density for r which is relatively noninformative, and has prior median equal to r^*.

Putting a prior on the parameter $\tilde{\lambda}$ for the transformed inefficiency error $\tilde{u}$ instead of on λ, which is the parameter for the original inefficiency error u, makes it apparent that efficiency measurement is not invariant with respect to scale reparameterizations of the model. Both of the above parameterizations will yield the same probability statements for each production category, but their respective efficiency measures, $\exp(-u_i)$ and $\exp(-\tilde{u}_i)$, will be different. This result is not surprising. We have observations only on categorical rankings of production, not on the absolute values of production. Consequently, we can only measure efficiency relative to a particular individual, or relative to that for a particular parameter setting, such as $\sigma_v^2 = 1$ or $\sigma_v^2 = 1/\gamma_{J-1}^2$. We return to this issue later.

Given our prior assumptions and the model in (22), the conditional posterior densities that can be used for Gibbs sampling are as follows. The conditional posterior densities for the transformed latent variables $\tilde{g}_i^*$ are independent truncated normal distributions

$$\left(\tilde{g}_i^* \mid \tilde{\gamma}, \tilde{\beta}, \sigma_v^2, \tilde{u}, \tilde{\lambda}, X, y\right) \sim N\left(x_i\tilde{\beta} - \tilde{u}_i, \sigma_v^2\right) I\left(\tilde{\gamma}_{y_i-1} < \tilde{g}_i^* < \tilde{\gamma}_{y_i}\right). \tag{23}$$

The conditional posterior density for $\tilde{\beta}$ is the normal distribution

$$\left(\tilde{\beta} \mid \tilde{\gamma}, \sigma_v^2, \tilde{u}, \tilde{\lambda}, \tilde{g}^*, X, y\right) \sim N\left((X'X)^{-1}X'(\tilde{g}^* + \tilde{u}), \sigma_v^2(X'X)^{-1}\right). \tag{24}$$

The inverted gamma conditional posterior density for σ_v^2 is

$$\left(\sigma_v^2 \mid \tilde{g}^*, \tilde{\beta}, \tilde{\gamma}, \tilde{u}, \tilde{\lambda}, X, y\right) \sim IG\left(\frac{N}{2} + a_v,\ b_v + \frac{1}{2}\left[(\tilde{g}^* - X\tilde{\beta} + \tilde{u})'(\tilde{g}^* - X\tilde{\beta} + \tilde{u})\right]\right). \tag{25}$$

When $J \geq 3$, the conditional posterior for $\tilde{\gamma}$ is required and is given by

$$p\left(\tilde{\gamma} \mid \tilde{\beta}, \sigma_v^2, \tilde{u}, \tilde{\lambda}, X, y\right) \propto \prod_{i=1}^{N} \left[\Phi\left(\frac{\tilde{\gamma}_{y_i} - x_i \tilde{\beta} + \tilde{u}_i}{\sigma_v} \right) - \Phi\left(\frac{\tilde{\gamma}_{y_i-1} - x_i \tilde{\beta} + \tilde{u}_i}{\sigma_v} \right) \right]. \tag{26}$$

A Dirichlet proposal density for $p\left(\tilde{\gamma} \mid \tilde{\beta}, \sigma_v^2, \tilde{u}, \tilde{\lambda}, X, y\right)$ is constructed, as described in the previous section. The probability of accepting the candidate draw $\tilde{\gamma}^{can}$ is $\min\{R,1\}$, where

$$R = \left\{ \prod_{i=1}^{N} \frac{\Phi\left(\left(\tilde{\gamma}_{y_i}^{can} - x_i \tilde{\beta} + \tilde{u}_i \right) / \sigma_v \right) - \Phi\left(\left(\tilde{\gamma}_{y_i-1}^{can} - x_i \tilde{\beta} + \tilde{u}_i \right) / \sigma_v \right)}{\Phi\left(\left(\tilde{\gamma}_{y_i} - x_i \tilde{\beta} + \tilde{u}_i \right) / \sigma_v \right) - \Phi\left(\left(\tilde{\gamma}_{y_i-1} - x_i \tilde{\beta} + \tilde{u}_i \right) / \sigma_v \right)} \right\} \left\{ \prod_{j=1}^{J-1} \left(\frac{q_j}{q_j^{can}} \right)^{\alpha_j n_j - 1} \right\}. \tag{27}$$

The conditional posterior density for the transformed inefficiency error $\tilde{u}_i$ is the truncated normal distribution

$$\left(\tilde{u}_i \mid \tilde{\lambda}, \sigma_v^2, \tilde{g}^*, \tilde{\beta}, \tilde{\gamma}, X, y\right) \sim N\left(x_i \tilde{\beta} - \tilde{g}_i^* - \tilde{\lambda}^{-1} \sigma_v^2,\ \sigma_v^2 \right) I\left(\tilde{u}_i \geq 0 \right). \tag{28}$$

Finally, the conditional posterior density for $\tilde{\lambda}$ is the inverted-gamma density

$$\left(\tilde{\lambda} \mid \tilde{\beta}, \tilde{\gamma}, \tilde{u}, \sigma_v^2, \tilde{g}^*, X, y\right) \sim IG\left(N+1,\ \sum_{i=1}^{N} \tilde{u}_i - \ln(r^*) \right). \tag{29}$$

If desired, values for the original untransformed parameters can be obtained at each iteration by dividing by σ_v.

2.4 Generalising the inefficiency term

In line with the literature for the stochastic frontier model with a continuous output variable, in this section we extend our model to allow the inefficiency term u_i to be related to explanatory variables. Of interest is whether individuals with some special characteristics are more likely to be more efficient than others. Our model specification and algorithm for Bayesian estimation follows that in Koop et al. (1997).

Suppose there are *m* variables, w_{ik} $(i = 1,\ 2,\ldots,N;\ k = 1,\ 2,\ldots,m)$, that impact on the efficiency of individuals, where $w_{i1} \equiv 1$ is a constant and w_{ik} $(k = 2,\ldots,m)$ are dummy variables representing individual characteristics. Continuous *w* variables can be introduced, but only at a cost of computation complexity. Define $W = (w_1,\ldots,w_N)'$ as an $N \times m$ matrix. Assume that u_i follows the exponential distribution:

$$u_i \sim \Gamma\left(1,\ \lambda_i^{-1}\right), \quad \lambda_i^{-1} = \prod_{k=1}^{m} \phi_k^{w_{ik}}, \tag{30}$$

where ϕ_k $(k = 1,\ldots,m)$ are unknown parameters. Since the mean of the inefficiency distribution λ_i is always positive, the ϕ_k should all be positive. If $\phi_k \equiv 1$ for $k = 2,\ldots,m$, $\lambda_i^{-1} \equiv \phi_1$ is a constant and the model collapses to the standard model in equation (5). Otherwise, the ϕ_k $(k = 2,\ldots,m)$ are to be estimated and the magnitude of a ϕ_k (in particular whether $\phi_k < 1$ or $\phi_k > 1$) determines whether the attribute w_k is a “bad” or “good” attribute in terms of its contribution to mean inefficiency. Since the *k*th term enters the product in (30) as $\phi_k^{w_k} = 1$ if $w_k = 0$, and $\phi_k^{w_k} = \phi_k$ if $w_k = 1$, individuals with attribute w_k will have a higher mean inefficiency λ_i if $\phi_k < 1$ (i.e., w_k is a “bad” attribute), and a lower mean inefficiency if $\phi_k > 1$ (w_k is a “good” attribute). Note that $\phi_k < 1$ does not mean that an individual with attribute w_k is definitely more inefficient than those without this characteristic, but rather that the former has an inefficiency u_i drawn from a distribution with a higher mean, assuming all other characteristics are the same.

For the reparameterized version that is used for estimation the transformations in (22) are adopted again except that $\tilde{\lambda} = \lambda/\gamma_{J-1}$ is replaced by

$$\tilde{\phi}_1 = \phi_1 \times \gamma_{J-1} \quad \text{and} \quad \tilde{\phi}_k = \phi_k \ (k = 2,\ldots,m), \tag{31}$$

from which we define

$$\tilde{\lambda}_i^{-1} = \prod_{k=1}^{m} \tilde{\phi}_k^{w_{ik}} \ . \tag{32}$$

Since $w_{i1} \equiv 1$, $\tilde{\lambda}_i = \lambda_i / \gamma_{J-1}$, which is in line with the previous section. Also, $\tilde{u}_i \sim \Gamma\left(1, \tilde{\lambda}_i^{-1}\right)$.

To specify a prior distribution for the new parameters $\tilde{\phi} = (\tilde{\phi}_1, \ldots, \tilde{\phi}_m)'$ we assume independent gamma priors where $\tilde{\phi}_k \sim \Gamma(a_k, b_k)$ and $p(\tilde{\phi}) = p(\tilde{\phi}_1) \cdots p(\tilde{\phi}_m)$, with a_k and b_k being hyperparameters. If none of the dummy variables has an impact on the efficiency distribution, i.e., $\tilde{\phi}_k \equiv 1$ ($k = 2, \ldots, m$), then $\tilde{\lambda}_i^{-1} \equiv \tilde{\phi}_1$; the model collapses to the standard model. This suggests the settings $a_1 = 1$ and $b_1 = -\ln(r^*)$ as discussed before. The other prior hyperparameters can be selected to yield a relatively noninformative prior.

The conditional posterior densities for the parameters $\tilde{\beta}$, $\tilde{\gamma}$ and σ_v^2, as well as that for the latent variable $\tilde{g}^*$, remain the same as in the standard model. The conditional posterior density for $\tilde{u}_i$ in (28) is only affected by changing $\tilde{\lambda}^{-1}$ to $\tilde{\lambda}_i^{-1}$ as defined in (32). The new parameters, $\tilde{\phi}_k$ ($k = 1, 2, \ldots, m$), can be drawn from

$$\left(\tilde{\phi}_k \mid \tilde{u}, \tilde{\phi}^{(-k)}, \tilde{g}^*, \tilde{\beta}, \tilde{\gamma}, \sigma_v^2, y, X, W\right) \sim \Gamma\left(a_k + \sum_{i=1}^{N} w_{ik},\ b_k + \sum_{i=1}^{N}\left(w_{ik}\tilde{u}_i \prod_{s \neq k} \tilde{\phi}_s^{w_{is}}\right)\right) \tag{33}$$

As before, the original parameters can be recovered from the transformed ones after each iteration of the Gibbs sampler.

3. Modelling with panel data

Panel data is commonly used in the stochastic frontier model for continuous variables. In classical analysis of such models, a relative efficiency measure can be obtained using a fixed effects specification, and an absolute efficiency measure is obtainable from a random effects specification. In Bayesian analysis, the difference between fixed and random effect models can be defined through the prior distribution for inefficiency u_i, but, otherwise, the two models are treated in the same way (Koop et al. 1997). Fixed effect models assume that the u_i's are drawn from fully separate distributions, while in random effect models the u_i's are linked by assuming they are drawn from distributions with a small number of unknown common parameter(s). In this section, we develop a random effects and time-invariant efficiency stochastic frontier model for ordinal outcomes with panel data.

3.1 Model specification

Assume we have a balanced panel data set for *N* individuals over *T* time periods. The methodology can be readily extended to an unbalanced panel data set, where not all *N* individuals have records for all *T* time periods. Indeed, our later application uses an unbalanced panel data set. The stochastic frontier model for ordinal outcomes can be written as

$$\ln y_{it}^* = f(x_{it}, \beta) + v_{it} - u_i \quad (i = 1, \ldots, N;\ t = 1, \ldots, T). \tag{34}$$

The latent variable y_{it}^* can be mapped to the observed values y_{it} as

$$y_{it} = \begin{cases} 0 & 0 < y_{it}^* \le 1 \\ 1 & 1 < y_{it}^* \le \mu_1 \\ & \vdots \\ j & \mu_{j-1} < y_{it}^* \le \mu_j \qquad j = 2, \ldots, J-1 \\ & \vdots \\ J & y_{it}^* > \mu_{J-1}. \end{cases} \tag{35}$$

In the special case where $f(x_{it}, \beta)$ is linear, (34) and (35) can be written as

$$
\begin{aligned}
&g_{it}^* = x_{it}\beta + v_{it} - u_i \qquad && (i = 1,\dots,N;\ t = 1,\dots,T) \\
&y_{it} = j, \ \text{ if and only if } \ \gamma_{j-1} < g_{it}^* \le \gamma_j \qquad && (j = 0,\dots,J)
\end{aligned}
\tag{36}
$$

where $g_{it}^* = \ln\left(y_{it}^*\right)$ and $\gamma_j = \ln\left(\mu_j\right)$, with $\gamma_{-1} = -\infty$, $\gamma_0 = 0$ and $\gamma_J = +\infty$. As in the case for cross-sectional data, two parameterizations are of interest: the traditional one where v_{it} are i.i.d. $N(0,1)$ and the one that is convenient for estimation where v_{it} are i.i.d. $N(0,\sigma_v^2)$ and there is one less unknown threshold. In line with the previous section, we distinguish between the two setups by using an over-tilde ($\sim$) for the second case. Thus for (36) we assume $v_{it} \sim$ i.i.d. $N\left(0,1\right)$. We again assume the u_i follow an exponential distribution, i.e., $u_i \sim$ i.i.d. $\Gamma\left(1, \lambda^{-1}\right)$. Extending the notation of the previous section, y_i, g_i^* and v_i are $T \times 1$ vectors and x_i is a $T \times (k+1)$ matrix, containing T observations for individual *i*. Because of the assumption of time-invariant efficiency, u_i is still a scalar. Further, we define $y = (y_1',\, y_1',\dots,y_N')'$, $g^* = (g_1^{*\prime},\, g_2^{*\prime},\dots,g_N^{*\prime})'$ and $v = (v_1',\, v_2',\dots,v_N')'$ as $NT \times 1$ vectors, and $X = (x_1',\, x_2',\dots,x_N')'$ as an $NT \times (k+1)$ matrix. We also let $u = (u_1,\, u_2,\dots,u_N)'$ and $\gamma = (\gamma_1,\, \gamma_2,\dots,\gamma_{J-1})'$.

For the transformed model the parameters, the latent variable and the two error terms are rescaled by dividing by the largest threshold parameter,

$$
\sigma_v = 1/\gamma_{J-1}, \ \ \tilde{\beta} = \beta\sigma_v, \ \ \tilde{\gamma} = \gamma\sigma_v, \ \ \tilde{g}^* = g^*\sigma_v, \ \ \tilde{v} = v\sigma_v, \ \ \tilde{u} = u\sigma_v, \ \ \tilde{\lambda} = \lambda\sigma_v \tag{37}
$$

from which we obtain $\tilde{v}_{it} \sim$ i.i.d. $N\left(0,\sigma_v^2\right)$ and $\tilde{u}_i \sim$ i.i.d. $\Gamma\left(1,\tilde{\lambda}^{-1}\right)$. For prior densities we use $p(\tilde{\beta},\tilde{\gamma}) \propto 1$, $p\left(\sigma_v^2\right) \propto 1/\sigma_v^2$, and $\tilde{\lambda} \sim IG\left(1,\ln(r^*)\right)$ where, as before, r^* is the prior median for the efficiency distribution. This prior specification is the same as that specified for the cross-sectional case except for the improper prior $p\left(\sigma_v^2\right) \propto 1/\sigma_v^2$ which can be used for the frontier model with panel data (Fernández et al 1997).

3.2 Conditional posteriors

Combining the various components, the joint posterior for the transformed parameters and latent variables is

$$p\left(\tilde{\beta}, \sigma_v^2, \tilde{\gamma}, \tilde{u}, \tilde{\lambda}, \tilde{g}^* \mid X, y\right) \propto \left[\prod_{i=1}^{N} \prod_{t=1}^{T} I\left(\tilde{\gamma}_{y_{it}-1} < \tilde{g}_{it}^* < \tilde{\gamma}_{y_{it}}\right)\right]\left[\prod_{i=1}^{N} I\left(\tilde{u}_i > 0\right)\right] \\ \times \frac{1}{\left(\sigma_v^2\right)^{NT/2+1} \tilde{\lambda}^{N+2}} \exp\left\{-\frac{1}{2\sigma_v^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{g}_{it}^* - x_{it}\tilde{\beta} + \tilde{u}_i\right)^2 - \frac{1}{\tilde{\lambda}}\left(\sum_{i=1}^{N} \tilde{u}_i - \ln(r^*)\right)\right\} \tag{38}$$

The conditional posterior densities for implementing the Gibbs sampler are:

$$\left(\tilde{g}_{it}^* \mid \tilde{\beta}, \sigma_v^2, \tilde{\gamma}, \tilde{u}, \tilde{\lambda}, X, y\right) \sim N\left(x_{it}\tilde{\beta} - \tilde{u}_i, \sigma_v^2\right) I\left(\tilde{\gamma}_{y_{it}-1} < \tilde{g}_{it}^* < \tilde{\gamma}_{y_{it}}\right) \tag{39}$$

$$\left(\tilde{\beta} \mid \sigma_v^2, \tilde{\gamma}, \tilde{u}, \tilde{\lambda}, \tilde{g}^*, X, y\right) \sim N\left(\left(X'X\right)^{-1} X'\left(\tilde{g}^* + (I_N \otimes \iota_T)\tilde{u}\right), \sigma_v^2 \left(X'X\right)^{-1}\right) \tag{40}$$

where I_N is an $N \times N$ identity matrix and ι_T is a $T \times 1$ vector of ones.

$$\left(\sigma_v^2 \mid \tilde{\beta}, \tilde{\gamma}, \tilde{u}, \tilde{\lambda}, \tilde{g}^*, X, y\right) \sim IG\left(\frac{NT}{2}, \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{g}_{it}^* - x_{it}\tilde{\beta} + \tilde{u}_i\right)^2\right) \tag{41}$$

For $J > 3$, $\tilde{\gamma}$ is drawn from the conditional posterior

$$p\left(\tilde{\gamma} \mid \tilde{\beta}, \sigma_v^2, \tilde{u}, \tilde{\lambda}, \tilde{g}^*, X, y\right) \propto \prod_{i=1}^{N} \prod_{t=1}^{T} \left\{\Phi\left(\frac{\tilde{\gamma}_{y_{it}} - x_{it}\tilde{\beta} + \tilde{u}_i}{\sigma_v}\right) - \Phi\left(\frac{\tilde{\gamma}_{y_{it}-1} - x_{it}\tilde{\beta} + \tilde{u}_i}{\sigma_v}\right)\right\}. \tag{42}$$

The threshold parameters can be drawn according to the M-H algorithm described in the previous section. The inefficiency term $\tilde{u}_i$ is drawn from

$$\left(\tilde{u}_i \mid \tilde{\beta}, \tilde{\gamma}, \sigma_v^2, \tilde{\lambda}, \tilde{g}^*, y, X\right) \sim N\left(\overline{x}_i \tilde{\beta} - \overline{\tilde{g}}_i^* - \left(T\tilde{\lambda}\right)^{-1} \sigma_v^2,\ T^{-1}\sigma_v^2\right) I\left(\tilde{u}_i \geq 0\right). \tag{43}$$

where $\overline{x}_i$ and $\overline{\tilde{g}}_i^*$ are the respective means of x_{it} and $\tilde{g}_{it}^*$ over the T observations for individual *i*. The conditional posterior for $\tilde{\lambda}$ is the same as for the cross-sectional case, i.e.,

$$\left(\tilde{\lambda} \mid \tilde{\beta}, \tilde{\gamma}, \tilde{u}, \sigma_v^2, X, y, \tilde{g}^*\right) \sim IG\left(N+1, \sum_{i=1}^{N} \tilde{u}_i - \ln(r^*)\right) \tag{44}$$

If we allow for explanatory variables to influence the distribution of inefficiency, all the other conditional posteriors will be the same as before except for the inefficiency error u and the new parameters ϕ. Assuming the same prior as in the cross-sectional case, the conditional posterior for $\tilde{u}_i$ from which draws can be taken is

$$\left(\tilde{u}_i \mid \tilde{\phi}, \tilde{g}^*, \tilde{\beta}, \tilde{\gamma}, \sigma_v^2, y, X, W\right) \sim N\left(\overline{x}_i \tilde{\beta} - \overline{\tilde{g}}_i^* - \left(T\tilde{\lambda}_i\right)^{-1} \sigma_v^2,\ T^{-1}\sigma_v^2\right) I\left(\tilde{u}_i \geq 0\right). \tag{45}$$

The conditional posterior distributions for the elements of $\tilde{\phi}$ are exactly the same as those for the case of cross-sectional data given in (33). Finally, draws for parameters in the original specification can be recovered by reversing the scale transformation in (37).

3.3 Test estimation with generated data

In this section, we generate two sets of artificial panel data to test the MCMC algorithms: one with and one without explanatory variables in the distribution of the inefficiency term. In Experiment 1, an unbalanced panel data set with $N = 10,000$ and maximum $T_i = 4$ is generated – a total of 30,031 observations – and no explanatory variables appear in the distribution of the inefficiency term u_i. The X variables are drawn from $X = (1, x_1, x_2, x_3)$, with two dummy variables $x_1 = I(U[0,1] > 0.45)$ and $x_3 = I(U[0,1] > 0.25)$, and one continuous variable $x_2 = \ln(U[0,100])$, where $U[a,b]$

denotes the uniform distribution on the interval $[a,b]$. Parameter values are set as $\beta = (1,\ 0.5,\ -0.15,\ 0.7)'$ and $\lambda^{-1} = 5$, with thresholds $\gamma = (0.6,\ 1.2,\ 2.0)'$, defining 5 production categories.. For MCMC estimation, the burn-in is taken as 2,000 iterations and the number of total recorded iterations after the burn-in is 10,000. The design of Experiment 2 is the same as Experiment 1, except that the total number of observations is 30,092 and there is an explanatory variable in the distribution of inefficiency. We set $W = (1,\ w_1),$ where $w_1 = I(U[0,1] > 0.25)$, and $\phi = (7,\ 0.5)'$.

Results from the two experiments are presented in Tables 1 and 2, respectively. These tables contain the true parameter values, the MCMC-estimated posterior means and standard deviations, and the 2.5 and 97.5 percentiles that define 95% credibility intervals. The results are very satisfying. We do not know the true values of the posterior means from a single sample, and so the table does not tell us the error from the MCMC-estimated means, but, nevertheless, obtaining MCMC-estimated means close to the true parameter values, and relatively small posterior standard deviations, is reassuring.

4. Quantities of interest

As in the ordered probit model, we are typically more interested in various functions of the parameters β and γ than in the parameters themselves. Two such functions are the probabilities of each ordered outcome and the marginal effects of changes in an x or a w on those probabilities. Also of interest are the efficiencies of individuals for given values of x and w.

4.1 Estimated probabilities and marginal effects

To obtain the posterior distributions of the outcome probabilities $\Pr(y = j)$, $j = 0,\ldots,J$, we begin by considering the probability of $y_s = j$ ($j = 0,\ldots,J$) for an out-of-sample individual s with observable covariates x_s and w_s and known parameters $\theta=(\beta,\gamma,\lambda)$; or with $\theta=(\beta,\gamma,\phi)$ if explanatory variables are allowed to influence the inefficiency.

$$\begin{aligned}\Pr\left(y_s = j \mid x_s, w_s, \theta\right) &= \Pr\left(\gamma_{j-1} < x_s\beta + v_s - u_s \le \gamma_j\right) \\ &= \Pr\left(\gamma_{j-1} - x_s\beta + u_s < v_s \le \gamma_j - x_s\beta + u_s\right) \\ &= \int_0^{+\infty}\left(\int_{\gamma_{j-1}-x_s\beta+u_s}^{\gamma_j-x_s\beta+u_s} p\left(v_s \mid \theta\right)dv_s\right)p\left(u_s \mid \theta\right)du_s \\ &= \int_0^{+\infty}\left(\Phi\left(\gamma_j - x_s\beta + u_s\right) - \Phi\left(\gamma_{j-1} - x_s\beta + u_s\right)\right)p\left(u_s \mid \theta\right)du_s\end{aligned} \tag{46}$$

Given a value for θ, this integral can be estimated by drawing a number of values of u_s, say 1,000, from its distribution $p\left(u_s \mid \theta\right)=\lambda^{-1}\exp\left(-u_s\lambda^{-1}\right)$, and then taking the average value of $\Phi\left(\gamma_j - x_s\beta + u_s\right)-\Phi\left(\gamma_{j-1} - x_s\beta + u_s\right)$ over all draws of u_s. If explanatory variables appear in the inefficiency distribution, we replace λ with $\lambda_s^{-1}=\prod_{k=1}^{m}\phi_k^{w_{sk}}$. Noting that

$$\Pr\left(y_s = j \mid x_s, w_s, y, X, W\right)=\int \Pr\left(y_s = j \mid x_s, w_s, \theta\right)p(\theta \mid y, X, W)\,d\theta,$$

to get draws from the posterior density for $\Pr\left(y_s = j \mid x_s, w_s\right)$, we repeat the above process for each draw of θ from the MCMC algorithm.

Consider now the marginal effects on the probabilities of a change in a continuous covariate, say x_{sk}, evaluated at settings $\left(x_s, w_s\right)$. From (46)

$$\frac{\partial \Pr\left(y_s = j \mid x_s, w_s, \theta\right)}{\partial x_{sk}}=\int_0^{+\infty}\left(\phi_{SN}\left(\gamma_j - x_s\beta + u_s\right)-\phi_{SN}\left(\gamma_{j-1} - x_s\beta + u_s\right)\right)\beta_k\, p\left(u_s \mid \theta\right)du_s \tag{47}$$

where $\phi_{SN}(\cdot)$ denotes the density of a standard normal random variable. Draws from the posterior density for $\partial \Pr\left(y_s = j \mid x_s, w_s\right) / \partial x_{sk}$ can be obtained using (47) in the same way that (46) was used to obtain draws from the posterior density for $\Pr\left(y_s = j \mid x_s, w_s\right)$.

For any binary variable *d* in *x* or *w*, the marginal effect is given by

$$\Pr\left(y = j \mid d = 1, x_s^*, w_s^*\right) - \Pr\left(y = j \mid d = 0, x_s^*, w_s^*\right) \tag{48}$$

where $\left(x_s^*, w_s^*\right)$ denotes the settings of all other variables at which the effect of *d* is evaluated. Draws from the posterior distributions of these quantities can be obtained by computing them for each MCMC draw of θ.

4.2 Efficiency measures

One of the main aims of traditional production frontier analysis is to evaluate and rank the efficiencies of all firms in the sample, given observed input and output levels of these firms. While this may be of interest in our micro level data application to the health production for individuals, we are more likely to be interested in predicting the efficiency of a particular out-of-sample individual whose health output has not been observed, or the average efficiency of out-of-sample individuals with particular characteristics. In this section, we introduce efficiency measures for an individual within the sample (where health output is observed), an individual out of the sample (where health output is not observed), and the average efficiency of out-of-sample individuals with particular characteristics w_s. We present results for the panel data model with explanatory variables *w* in the distribution of the inefficiency term. For cross-sectional data or a simpler version of the model, similar results can be obtained.

Before turning to these results, we discuss limits to efficiency measurement that are a consequence of having ordered categorical data rather than a continuous fully-observed output variable. When the original model was transformed to a model with one less unknown threshold and an extra variance parameter $\left(\sigma_v^2 \neq 1\right)$, we mentioned that efficiency measurement is not invariant with respect to scale transformations of that nature. To further appreciate this fact and to explore ways of presenting information on relative efficiency, consider the panel-data frontier model

$$
\begin{aligned}
g_{it}^* = \ln y_{it}^* = x_{it}\beta + v_{it} - u_i, \qquad & v_{it} \sim N(0, \sigma_v^2), \\
u_i \sim \Gamma\left(1, \lambda_i^{-1}\right), \qquad & \lambda_i^{-1} = \phi_1 \prod_{k=2}^{m} \phi_k^{w_{ik}}
\end{aligned}
\tag{49}
$$

where the category $y_i = j$ is observed when

$$
\gamma_{j-1} < g_i^* \leq \gamma_j \qquad j = 0,1,\ldots,J\,, \quad \text{with } \gamma_{-1} = -\infty,\ \gamma_0 = 0 \text{ and } \gamma_J = +\infty. \tag{50}
$$

When g_i^* is observed, only the equations in (49) are considered and all parameters are identified; we have the traditional frontier model where efficiency measurement is well defined. When y_i but not g_i^* is observed, we consider both (49) and (50) and not all of the parameters are identified. The two ways of achieving identification that we have considered are setting $\sigma_v^2 = 1$ or $\gamma_{J-1} = 1$. We now ask what would be the effect of these types of restrictions on efficiency measurement if g_i^* was observed? They imply we are considering efficiency defined by a transformed error of the form u_i/σ_v or u_i/γ_{J-1}. Consider, for example, the error u_i/σ_v. From (49), its distribution is

$$
\left(u_i/\sigma_v\right) \sim \Gamma\left(1, \lambda_i^{-1}/\sigma_v\right) \quad \text{where} \quad \lambda_i^{-1}/\sigma_v = \left(\phi_1/\sigma_v\right) \prod_{k=2}^{m} \phi_k^{w_{ik}} \tag{51}
$$

Thus, failure to identify σ_v (failure to observe g_i^*), and setting $\sigma_v^2 = 1$ to overcome this problem, means we are estimating an inefficiency distribution with first parameter $\phi_1^* = \left(\phi_1/\sigma_v\right)$ when the correct inefficiency distribution parameter should be ϕ_1. Because we cannot retrieve ϕ_1 from ϕ_1^*, we cannot estimate the absolute level of efficiency or inefficiency implied by ϕ_1. However, we can define an arbitrary absolute level of efficiency (and inefficiency) by setting σ_v^2 equal to a specific value, and then examining how efficiencies change relative to that level for different settings of $w_{ik}, k = 2,\ldots,m$. In our application, we took $w_{ik} = 0$ for $k = 2,\ldots,m$ as our reference setting, defined a level of efficiency for that setting, and then examined the efficiency implications of other w_{ik} values. What we mean by "a level of efficiency" is made more precise in the application, after we have considered the various efficiency measures.

4.2.1 Efficiency measure for the i^{th} individual in the sample data set

As defined in Section 2.1, the efficiency of the i^{th} individual is $r_i = e^{-u_i}$ $(0 < r_i \le 1)$. To assess the efficiency of this sample individual given the observed data, we are interested in the posterior density function, $p\left(r_i \mid X, W, y\right)$ and its mean and variance. We derive expressions for these quantities and the other efficiency measures under the assumption that $\sigma_v^2 = 1$. If another setting of σ_v^2 is used, u_i needs to be scaled accordingly.

The inefficiency term u_i conditional on $\Theta = \left(\beta, \gamma, \phi, g^*\right)$, X and W follows the truncated normal distribution

$$\left(u_i \mid \Theta, X, W\right) \propto N\left(\overline{x}_i\beta - \overline{g}_i^* - (T\lambda_i)^{-1}, T^{-1}\right) I\left(u_i \ge 0\right), \tag{52}$$

where $\bar{x}_i$ and $\bar{g}_i^*$ are the respective means of x_{it} and g_{it}^* over T observations for the ith individual, and $\lambda_i^{-1} = \prod_{k=1}^{m} \phi_k^{w_{ik}}$. This distribution is obtained from (45) after transforming back to the original parameters. Its density $p(u_i \mid \Theta, X, W)$ can be written as

$$p\left(u_i \mid \Theta, X, W\right) = \frac{I\left(u_i \geq 0\right)}{\sqrt{2\pi}\sigma_{u_i}\Phi\left(\mu_{u_i} / \sigma_{u_i}\right)} \exp\left\{-\frac{1}{2}\left(\frac{u_i - \mu_{u_i}}{\sigma_{u_i}}\right)^2\right\}, \tag{53}$$

where $\mu_{u_i} = \bar{x}_i\beta - \bar{g}_i^* - \left(T\lambda_i\right)^{-1}$, and $\sigma_{u_i} = 1/T$. Using a transformation of variables, the conditional posterior density for efficiency of the *i*-th individual is

$$p\left(r_i \mid \Theta, X, W\right) = \frac{I(0 < r_i \leq 1)}{\sqrt{2\pi}\,\sigma_{u_i} r_i \,\Phi\left(\mu_{u_i} / \sigma_{u_i}\right)} \exp\left\{-\frac{1}{2}\left(\frac{-\ln r_i - \mu_{u_i}}{\sigma_{u_i}}\right)^2\right\}. \tag{54}$$

Thus, the unconditional posterior density for a within-sample individual's efficiency is given by

$$p\left(r_i \mid X, W, y\right) = \int p\left(r_i \mid \Theta, X, W\right) p\left(\Theta \mid X, W, y\right) d\Theta \tag{55}$$

which can be estimated using

$$\hat{p}\left(r_i \mid X, W, y\right) = \frac{1}{M}\sum_{n=1}^{M} f\left(r_i \mid \Theta^{(n)}, X, W\right), \tag{56}$$

where M is the total number of recorded iterations in the MCMC estimation, and $\Theta^{(n)}$ is the value of Θ generated in the n^{th} iteration. The average in (56) is carried out for a grid of values of r_i in the (0,1) interval. Alternatively, the density can be estimated directly using the MCMC draws of e^{u_i} .

To obtain the mean and variance of $p\left(r_i \mid X,W,y\right)$ the following result is useful. Let z be a non-negative truncated normal random variable, i.e. $z \sim N\left(\mu_z,\sigma_z\right)I(z\geq 0)$, then

$$E\left(\exp(-qz)\right)=\frac{\Phi\left(\mu_z/\sigma_z-q\sigma_z\right)}{\Phi\left(\mu_z/\sigma_z\right)}\exp\left(\frac{q^2\sigma_z^2}{2}-q\mu_z\right). \tag{57}$$

Thus, the first and second moments for the posterior density for r_i conditional on Θ are

$$\begin{aligned} E\left(r_i \mid \Theta,X,W\right) &= E\left(\exp\left(-u_i\right)\mid \Theta,X,W\right) \\ &= \frac{\Phi\left(\mu_{u_i}/\sigma_{u_i}-\sigma_{u_i}\right)}{\Phi\left(\mu_{u_i}/\sigma_{u_i}\right)}\exp\left(\frac{\sigma_{u_i}^2}{2}-\mu_{u_i}\right), \end{aligned} \tag{58}$$

and

$$\begin{aligned} E\left(r_i^2 \mid \Theta,X,W\right) &= E\left(\exp\left(-2u_i\right)\mid \Theta,X,W\right) \\ &= \frac{\Phi\left(\mu_{u_i}/\sigma_{u_i}-2\sigma_{u_i}\right)}{\Phi\left(\mu_{u_i}/\sigma_{u_i}\right)}\exp\left(2\sigma_{u_i}^2-2\mu_{u_i}\right). \end{aligned} \tag{59}$$

The unconditional posterior moments $E\left(r_i \mid X,W,y\right)$ and $E\left(r_i^2 \mid X,W,y\right)$ can be estimated by averaging (58) and (59) over the MCMC draws for Θ, and an estimate for the posterior variance of r_i is calculated from these quantities.

4.2.2 Efficiency measure for an out-of-sample individual

Suppose that interest centers on the efficiency of an out-of-sample individual with characteristics w_s and corresponding inefficiency error u_s that is a drawing from an exponential distribution with density

$$p\left(u_s \mid w_s,\phi\right)=\lambda_s^{-1}\exp\left(-u_s\lambda_s^{-1}\right)I\left(u_s\geq 0\right),\ \text{ where }\ \lambda_s^{-1}=\prod_{k=1}^{m}\phi_k^{w_{sk}} \tag{60}$$

From (60), the density function for the efficiency of this individual, $r_s = \exp(-u_s)$, is

$$p\left(r_s \mid w_s, \phi\right) = \lambda_s^{-1} r_s^{\left(\lambda_s^{-1}-1\right)} I\left(0 < r_s \le 1\right). \tag{61}$$

Its first and second moments are $E\left(r_s \mid w_s, \phi\right) = \left(\lambda_s + 1\right)^{-1}$ and $E\left(r_s^2 \mid w_s, \phi\right) = \left(2\lambda_s + 1\right)^{-1}$, respectively. These results are different from those for a within-sample individual because we no longer condition on the person's *y* and *x* values which are not observed. However, the sample values *y*, *X* and *W* provide information on ϕ through its posterior density which is used to obtain the Bayesian predictive density

$$p\left(r_s \mid w_s, X, W, y\right) = \int p\left(r_s \mid w_s, \phi\right) p\left(\phi \mid X, W, y\right) d\phi \tag{62}$$

We estimate this density by first computing $\lambda_{s(n)}^{-1} = \prod_{k=1}^{m} \phi_{(n)k}^{w_{sk}}$ for each MCMC draw $\phi_{(n)}$, $n = 1, 2, \ldots, M$, and then averaging (61) over the *M* values of $\lambda_{s(n)}^{-1}$ for a grid of values of r_s in the interval (0,1). Similarly, estimates of its moments $E\left(r_s \mid w_s, X, W, y\right)$ and $E\left(r_s^2 \mid w_s, X, W, y\right)$ are obtained by averaging $\left(\lambda_{s(n)} + 1\right)^{-1}$ and $\left(2\lambda_{s(n)} + 1\right)^{-1}$ over $\lambda_{s(n)}$.

The density $p\left(r_s \mid w_s, X, W, y\right)$ and its mean and variance are used to provide information about the efficiency of a randomly selected individual from the population with specific attributes w_s. It includes variation from not knowing the parameters ϕ, and from the random selection of an individual from the population. In our application it provides an answer to a question such as: If an individual who drinks heavily, smokes and never exercises is drawn randomly from the population, what are the likely values of that person's health efficiency?

4.2.3 Average efficiency of out-of-sample individuals

A third efficiency measure likely to be of interest is the “average” performance of out-of-sample individuals with particular attributes w_s. For example, we might be interested in the average efficiency of all individuals who drink heavily, smoke and never exercise. What we require is the posterior density for $E\left(r_s \mid w_s, \phi\right) = \left(\lambda_s + 1\right)^{-1}$ which can be estimated from the MCMC draws $\lambda_{s(n)}$. In this case variation comes only from the uncertainty in ϕ. The posterior mean of $\left(\lambda_s + 1\right)^{-1}$ is the same as the mean of the predictive density $p\left(r_s \mid w_s, X, W, y\right)$, but the posterior variance of $\left(\lambda_s + 1\right)^{-1}$ will be much smaller than that of $p\left(r_s \mid w_s, X, W, y\right)$ because it does not include the randomness of selecting a particular individual.

5. An application to health production of individuals

5.1 Data and specification of variables

The data used in this application are from the first fives waves of the Australian Household, Income and Labour Dynamics in Australia (HILDA) surveys conducted from 2001 to 2005. These surveys utilise a multi-stage sampling approach stratified by state and part-of-state. The HILDA data set is a nationally representative longitudinal one with broad information on individual and household characteristics over time. It also supplies a large amount of information on health status and health related behaviour, as well as demographic, socioeconomic, geographic and lifestyle characteristics of individuals. Some information is collected by face-to-face interview, while some is collected by a self-completed questionnaire which is collected at a later date or returned by post. Most of the lifestyle factors are asked in the self-completed questionnaire, and hence there are a relatively larger number of missing values on these variables. For this study, our sample is restricted to those 18 years or older, which involves 65,449 records. After removal of missing values, a sample of 53,164 records for 15,450 individuals was used. It is an unbalanced panel data set.

Definitions of all variables are given in Appendix A. The health status of individuals, the output variable in our production frontier, is presented in the form of self-reported health. It is collected through the question "*In general, would you say your health is: Excellent (4), Very good (3), Good (2), Fair (1), or Poor (0)*". Of the pooled sample of 53,164 records, only 3.41% reported poor health status, 13.97%, 34.44%, 35.60% and 12.58% stated fair, good, very good and excellent health, respectively.

The stochastic frontier model has two sets of covariates, *X* and *W*. In the traditional production frontier models (Battese and Coelli 1995), *X* represents the inputs of production and *W* relates to firm characteristics that may influence the efficiency of production. In the health production frontier specified here, demographic, socioeconomic and geographic factors, as well as specific health conditions are used for the *X* covariates (Desai 1987; Contoyannis and Jones 2004). We assume personal lifestyle behaviour, given *X*, influences the efficiency of health production via the *W* covariates. As shown in Appendix A, the *X* variables include gender, marriage status, natural logarithm of age and its square, country of birth, education level, long term chronic health conditions, remoteness of residency region, work status, and home ownership. For the *W* covariates, we use exercise level, smoking status, alcohol consumption and a social net work measure. As in our panel data model *W* is individual specific but time invariant. It is defined as the overall lifestyle behaviour over the five waves. Although we do observe some lifestyle changes over time, the changes are small over the five year period. Detailed definitions of the variables in *W* can also be found in Appendix A.

Descriptive statistics for observed health status by individual characteristics are presented in Appendix B. Individuals with higher education levels produce higher health status. Less than 2% of people with a higher degree stated poor health status, while this percentage was more than 5% for those with less than 12 years education. Full-time students and employed people were more likely to produce good health compared with those retired or not in labour force. Cultural origin of individuals was another crucial factor. Australian aboriginals report the worst health status; more than 7% of them had poor health and only 11% reported excellent health status. As expected, over 12% of those with a long term health condition reported poor health for the present period and only 2.5% of them reported excellent health status. Conversely, less than 1% of persons without a long term condition produced poor health and as high as 15.3% of them produced the highest level of health. The descriptive statistics also show that married people and females are better producers than their counterparts.

5.2 Results for estimated parameters

The Gibbs sampling algorithm utilized the transformed model where $\tilde{\gamma}_{J-1} = 1$ and $\sigma_v^2 \neq 1$, but for the presentation of results the parameters were transformed back to the original specification. For the prior specification on $\tilde{\phi}$, we set $r^* = 0.7$ and $a_k = 2$, $b_k = 2$, $k = 2, 3, \ldots, m$, to yield relatively noninformative priors. Improper noninformative priors were used for the other parameters. The burn-in period was taken as 2,000 iterations and the number of recorded iterations after the burn-in was 10,000. For assessing mixing performance, we adopted the simulation inefficiency factor (SIF) (Kim, et al. 1998), and plotted the MCMC sample paths of some selected typical parameters (Figure 1), and the autocorrelation functions of these sample paths (Figure 2). These graphs suggest that the sample paths are reasonably well mixed.

The posterior estimates for the parameters are summarized in Table 3. Looking first at the SIF values, we find they are all less than 60 and most of them are lower than 10, a quite strong indication of the convergence of the sampler. The estimated parameters ϕ all have 95% credibility intervals that do not include 1, implying all explanatory variables adopted in this model have significant impacts on the distribution of efficiency. In particular, doing exercise has a positive impact on the efficiency of health production, and the more exercise the individual does, the more efficient the individual is likely to be in health production. Unsurprisingly, never smoking also has a positive impact on efficiency. For alcohol consumption, individuals never drinking or those drinking too much are more likely to have low efficiency in health production relative to the base group of moderate drinkers. Finally, those who never feel lonely or are lonely only sometimes, have a higher mean efficiency compared to those who always feel lonely.

Because the parameters β represent the effects of X variables on the latent health variable, they are not invariant with respect to scale transformations, and their magnitudes have no direct meaning. However, they do indicate the direction and the ranking of the effects of the X variables. As shown in Table 3, the significant variables all have the expected signs. Controlling for all the other factors, females are more likely to produce good health status, and education is a positive input. Compared to full-time employed persons, full time students are doing better, while people fully retired or not in labour force are worse. There is no significant difference between full time employees and part time employees or unemployed persons. People owning a house or having a mortgage are more likely to produce good health status compared with their renting counterparts.

5.3 Marginal effect of X

For the marginal effects of the *X* variables on health status probabilities we first note there are only two continuous variables in *X*, both relating to age, while the rest are dummies. As discussed in Section 4.1, for each dummy variable the marginal effect is calculated as the difference between the probabilities when the dummy is turned on and off, with all other variables held at their sample means. For the two age-related continuous variables, instead of deriving $\partial\left(\Pr(y=j)\right)/\partial x$, the posteriors for $\Pr(y=j)$, $j=0,\ldots,4$, were graphed against age to give a complete picture of the impact of age over the life cycle.

Summary statistics for the posterior distributions for the marginal effects of the dummy variables are presented in Table 4, and the complete posterior densities for the marginal effects of some selected covariates on $\Pr(y=4)$, the probability of having "excellent" health status, are plotted in Figure 3. All the significant marginal effects have the expected signs. Consider first the basic demographic factors, gender, marital status and cultural origin. The posterior mean for the marginal effect of being a male on poor health is 0.19%, which means males are 0.19% more likely to produce poor health than females, while males are 0.83% less likely to have excellent health, holding other factors constant. And from the 95% credibility interval in Table 4 and the density of the marginal effect of gender on $\Pr(y=4)$ in Figure 3, there is a 95% probability that males are from 1.4% to 0.3% less likely to report excellent health. A married or partnered person is -0.83% to 0.24% more likely to be in excellent health status with 95% probability, with zero included in the credibility interval. Compared with Australian born non-aboriginals, people born in other main English speaking countries are 2.45% more likely to have excellent health status and 0.49% less likely to have poor health. While 0 is included in the 95% credibility intervals for the marginal effects of being Australian aboriginal on poor and excellent health outputs, the aboriginals are 3.6% to 0.0% less likely to report excellent health with 95% probability. People born in other countries are found to be 1.64% to 0.03% less likely to have excellent health status. Education level, as an important proxy of socioeconomic status, is a positive input in health production. Compared with those who have less than year 12 education, people with a higher degree are 1.36% less likely to report poor health output and 6.75% more likely to have excellent health status. These two numbers are 0.33% and 1.22% for persons with a diploma and 0.75% and 3.08% for those with year 12 education. The value 0 is not included in the 95% credibility intervals nor in the range of probability density functions for the marginal effects of the education dummies, providing strong evidence that the effects of education levels on health output are significant. House ownership also has a significant positive impact on having good health. Having a long term health condition will increase the probability of having poor health status by 2.76% and decrease the probability of excellent health status by 8.31%; the 95% credibility intervals for these two effects are far from 0. Major work activity is another important input in health production. Compared with the full time employed, full time students are more likely to produce good health, with a 0.42% less probability of having poor health and a 2.45% greater chance of excellent health. There appears to be no significant difference in health production between full time employees and other people in the labour force (i.e. part time employees and unemployed). Not surprisingly, people fully retired and not in the labour market are worse producers of health. They are 3.42% and 3.63% less likely to have excellent health, respectively.

Figure 4 presents the effects of age on the health probabilities with other exogenous variables set at their sample means. In Figure 4(a) the probability of poor health status $(y = 0)$ increases monotonically with age. It increases from 3.38% for those aged 18, to

8.03% for those aged 90, with the increase being slower before the age of 45 and steeper after 70. Interestingly, for the most popular choice of health category, that of "very good" (Figure 4(d)), the probability decreases from nearly 50% for the young to less than 30% for the old, with the rate of decrease much sharper after the age of 50. Finally, the probability of reporting "excellent" health decreases from 17% to 2% as age increases. The rate of decrease is slow before 23 years old then increases between 23 to 40 years, before slowing down again to 90 years old with a much smaller variance. From the above analysis, it can be concluded that the impact of age on the probabilities of health status is different at different age levels; reporting single measures for the marginal effects of age, evaluated, say, at the sample mean value of age, would conceal a great deal of information.

5.4 Efficiency measures

As discussed in Section 4.2, having output defined in terms of an ordered categorical variable instead of a continuous one means we cannot obtain an absolute measure of efficiency. However, we can choose a reference group with particular characteristics, say w_{ref}, set σ_v to define a posterior mean efficiency for that group, and then compare the efficiency distributions for other settings of *w*. It is convenient to choose as a reference group that where $w_1 = 1$, and $w_2 = w_3 = \cdots = w_8 = 0$. This group consists of those who never exercise, smoke, drink a moderate amount of alcohol, and always feel lonely. With the exception of the alcohol variable, these are the characteristics that lead to the worst level of efficiency. Perhaps surprisingly, the setting of the alcohol variable that leads to the least efficiency is "no alcohol"; that which leads to the greatest efficiency is a moderate amount of alcohol.

After some experimentation, we set $\sigma_v = 1/3$ as a convenient value that led to a mean efficiency of approximately 0.5 for an out-of-sample reference person. For other settings of *w* we chose a "best" individual (one with a high level of exercise, does not smoke, has a moderate level of alcohol, and never feels lonely), a 'worst" individual (never exercises, smokes, drinks no alcohol, and always feels lonely), and an "average" individual (the *w* variables are set at their sample means). We focus on efficiency for out-of-sample individuals, both the predictive density for efficiency of a randomly selected individual, and the posterior density for mean efficiency of individuals, with best, worst and average characteristics. As discussed in Section 4.2, the efficiency measure for an individual from the population with attributes *w* is likely to be more interesting than the efficiency of one within-sample individual chosen from a sample of over 15,000.

Table 5 shows the means and standard deviations of the predictive densities for efficiency and the posterior densities for mean efficiency for the different settings. The reference category is also included. Compared to the posterior mean of efficiency for the reference group of 0.48, the means of the worst, average and best groups are 0.43, 0.67 and 0.78, respectively. The standard deviations for the predictive densities for a randomly selected individual are much greater than those for the means of all individuals in the population. This observation is particularly evident from Figures 5 and 6 where the predictive and posterior densities are graphed.

Looking first at Figure 5, we see that an individual with the worst characteristics can have an efficiency anywhere in the (0,1) range, although efficiencies closer to zero are more probable. Someone with the best characteristics can have an efficiency anywhere between 0.1 and 1, but most of the probability is at the right end of the density with a low probability of an efficiency less than 0.4. The density for an average individual is almost linear, rising steadily over the range 0 to 1. Relative to those in the best category, there is a larger probability of a low efficiency and a smaller probability of a higher efficiency.

Moving to the posterior densities for mean efficiency given in Figure 6, we find that the densities for the different settings no longer overlap. Efficiency is measured with greater precision and the densities appear normally distributed as one would expect from parameter estimation in a large sample. The greater precision of the best category relative to that of the worst category reflects the larger number of observations in that category. The average category, which has the greatest precision, is an artificial one where variables are set equal to their sample means. If the setting for σ_v^2 is changed, and hence the reference setting changes, the location of each of the densities changes, but their relative positions remain the same.

6. Summary

We present a stochastic frontier model for discrete ordinal outcomes for both cross-sectional and panel data. The model is a meaningful extension of the stochastic frontier model with a continuous output variable. More generally, with the increasing use of unit record data in social science, in which much information is in the form of discrete ordinal data, this model has potential applications in other fields. Gibbs sampling with data augmentation is adopted as the posterior simulator, and a reparameterization algorithm is introduced to improve the simulation performance of the threshold parameters. The algorithm worked well when applied to test models with generated data. Posterior distributions for quantities of interest, including probabilities of outcome status, the marginal effects of inputs on output status, and efficiency measures, are also presented.

The model is applied to health production analysis using panel data from the HILDA survey. The basic demographic variables, education level, health stock and major activity, are taken as health production inputs and we allow for lifestyle factors, such as exercise level, alcohol consumption, smoking habits, and social network, to impact on the efficiency distribution. The marginal effects of inputs on the probabilities of health status are used to present the impact of inputs on health production output. Results on the impacts of all the health production inputs are consistent with expectations. Significant impacts of the lifestyle factors on health production efficiency are also found.

Our extension of the stochastic frontier model to discrete ordered dependent variables is based on traditional stochastic frontier models, in the spirit of Battese and Coelli (1988, 1995), Kumbhakar, et al. (1991), and Koop et al. (1997). Greene (2004, 2005) discusses the issue of distinguishing between individual heterogeneity and inefficiency in stochastic frontier analysis. He examined several extensions to the commonly used stochastic frontier model specifications for panel data to allow for more flexibility in accommodating firm heterogeneity while preserving the inefficiency measurement feature of the frontier models. These include the 'true' fixed and random effect models that have both the traditional fixed/random individual-specific term, as typically used in panel data linear regression models, as well as the one-sided inefficiency error term. He also presented random coefficient and latent class versions of the stochastic frontier model for isolating individual heterogeneity. In Greene's (2004, 2005) context, our model has allowed for individual heterogeneity to affect both the production function and the inefficiency term via observable time-invariant characteristics; it does not separately identify individual heterogeneity and inefficiency due to unobservable factors. Allowing for these possibilities is a potential avenue for future research.

Acknowledgements

This research was supported by an Australian Research Council Discovery Grant. Zhao thanks Don Poskitt for useful discussions at an early stage of the work.

Table 1

Results for experiment 1

	True	mean	St. D	2.50%	97.50%
β					
	1	1.008	0.033	0.943	1.073
	0.5	0.492	0.013	0.466	0.518
	-0.15	-0.150	0.006	-0.162	-0.137
	0.7	0.700	0.015	0.671	0.729
λ^{-1}					
	5	4.824	0.389	4.166	5.676
γ					
	0.6	0.592	0.005	0.583	0.601
	1.2	1.208	0.010	1.186	1.226
	2	2.004	0.013	1.979	2.029

Note: Mean and St.D. refer to the MCMC-estimated posterior mean and standard deviation of the parameter. 2.5% refers to the lower value of the 95% credibility interval, and 97.5% refers to the upper value of the 95% credibility interval.

Table 2

Results for experiment 2

	True	mean	St. D	2.50%	97.50%
β					
	1	1.024	0.032	0.962	1.086
	0.5	0.495	0.013	0.469	0.521
	-0.15	-0.151	0.006	-0.164	-0.139
	0.7	0.705	0.015	0.675	0.733
ϕ					
	7	6.723	0.845	5.282	8.636
	0.5	0.521	0.050	0.422	0.619
γ					
	0.6	0.591	0.005	0.579	0.599
	1.2	1.188	0.007	1.173	1.203
	2	2.017	0.013	1.992	2.041

Note: Mean and St.D. refer to the MCMC-estimated posterior mean and standard deviation of the parameter. 2.5% refers to the lower value of the 95% credibility interval, and 97.5% refers to the upper value of the 95% credibility interval.

Table 3

Posterior information on parameters for the stochastic frontier model for health production

	Mean	2.50%	97.50%	SIF
β				
ONE	1.476	-0.277	3.141	5.468
MALE	-0.059	-0.100	-0.019	8.284
MARRIAGE	-0.021	-0.059	0.017	6.794
LNAGE	2.648	1.727	3.620	5.695
LNAGE2	-0.464	-0.598	-0.337	5.764
AUSABO	-0.153	-0.306	0.001	9.334
MAINENG	0.163	0.102	0.223	7.502
OTHERC	-0.063	-0.125	-0.002	7.125
DEGREE	0.452	0.399	0.505	8.715
DIPLOMA	0.098	0.051	0.146	9.176
YEAR12	0.230	0.172	0.290	8.594
LONGC1	-0.748	-0.787	-0.710	6.196
INNER	0.001	-0.039	0.039	5.647
OUTER	-0.068	-0.122	-0.015	7.533
REMOTE	-0.021	-0.129	0.089	4.788
STUDENT	0.151	0.078	0.223	4.102
PARTTIME	-0.036	-0.075	0.003	5.181
UNEMP	-0.074	-0.149	0.002	3.938
RETD	-0.258	-0.318	-0.197	6.464
NOTINLAB	-0.276	-0.324	-0.228	5.814
HOUSEOWN	0.161	0.125	0.197	4.862
ϕ				
ONE	0.304	0.276	0.337	58.336
DOEX	1.457	1.334	1.583	46.757
ALDOEX	2.063	1.882	2.249	44.239
NOSM	1.237	1.191	1.285	5.346
NOA	0.819	0.771	0.867	5.878
HIGHA	0.930	0.882	0.981	3.528
LONELY0	1.507	1.422	1.597	14.945
LONELY1	1.128	1.062	1.198	13.845
γ				
γ_1	1.691	1.673	1.708	6.792
γ_2	3.489	3.453	3.524	6.792
γ_3	5.356	5.302	5.410	6.792

Note: Mean refers to the posterior mean of the parameter. 2.5% refers to the lower value of the 95% credibility interval, and 97.5% refers to the upper value of the 95% credibility interval.

Table 4

Posterior summary statistics for marginal effects of dummy variables (%) [a]

	Pr($y=0$)			Pr($y=1$)			Pr($y=2$)			Pr($y=3$)			Pr($y=4$)		
	Mean (St. D)	**2.50%**	**97.50%**	**Mean (St. D)**	**2.50%**	**97.50%**	**Mean (St. D)**	**2.50%**	**97.50%**	**Mean (St. D)**	**2.50%**	**97.50%**	**Mean (St. D)**	**2.50%**	**97.50%**
GENDER	0.19(0.07)	0.06	0.31	0.41(0.14)	0.13	0.69	0.94(0.33)	0.30	1.59	-0.71(0.25)	-1.20	-0.22	-0.83(0.29)	-1.39	-0.26
MARRIAGE	0.06(0.06)	-0.06	0.19	0.14(0.14)	-0.12	0.41	0.33(0.31)	-0.28	0.94	-0.24(0.23)	-0.70	0.21	-0.29(0.28)	-0.83	0.24
AUSABO	0.52(0.28)	-0.00 [b]	1.08	1.13(0.60)	-0.01	2.34	2.41(1.23)	-0.02	4.74	-2.14(1.18)	-4.55	0.02	-1.91(0.93)	-3.59	0.02
MAINENG	-0.49(0.09)	-0.66	-0.31	-1.07(0.20)	-1.45	-0.68	-2.57(0.49)	-3.52	-1.61	1.68(0.29)	1.10	2.22	2.45(0.49)	1.49	3.42
OTHERC	0.20(0.10)	0.01	0.41	0.45(0.22)	0.01	0.90	0.99(0.49)	0.03	1.98	-0.81(0.41)	-1.65	-0.03	-0.84(0.40)	-1.64	-0.03
DEGREE	-1.36(0.09)	-1.53	-1.19	-2.99(0.18)	-3.34	-2.63	-7.09(0.44)	-7.94	-6.25	4.69(0.29)	4.11	5.26	6.75(0.44)	5.91	7.59
DIPLOMA	-0.33(0.08)	-0.50	-0.17	-0.73(0.18)	-1.08	-0.37	-1.55(0.38)	-2.29	-0.80	1.38(0.34)	0.71	2.05	1.22(0.30)	0.64	1.82
YEAR12	-0.75(0.10)	-0.94	-0.56	-1.63(0.21)	-2.05	-1.22	-3.63(0.47)	-4.57	-2.70	2.93(0.36)	2.21	3.66	3.08(0.42)	2.26	3.91
LONGC1	2.76(0.11)	2.55	2.97	5.97(0.21)	5.55	6.40	11.21(0.32)	10.59	11.85	-11.63(0.44)	-12.50	-10.77	-8.31(0.19)	-8.69	-7.95
INNER	0.00(0.06)	-0.12	0.13	0.00(0.14)	-0.26	0.27	-0.01(0.31)	-0.61	0.62	0.01(0.23)	-0.47	0.45	0.01(0.28)	-0.55	0.55
OUTER	0.22(0.09)	0.05	0.40	0.48(0.19)	0.10	0.87	1.08(0.43)	0.23	1.93	-0.86(0.35)	-1.57	-0.18	-0.92(0.36)	-1.63	-0.20
REMOTE	0.07(0.18)	-0.27	0.42	0.15(0.38)	-0.59	0.92	0.34(0.87)	-1.40	2.02	-0.28(0.67)	-1.67	0.97	-0.28(0.76)	-1.70	1.29
STUDENT	-0.42(0.10)	-0.62	-0.22	-0.93(0.22)	-1.36	-0.50	-2.36(0.57)	-3.48	-1.23	1.27(0.27)	0.71	1.77	2.45(0.63)	1.24	3.72
PARTTIME	0.11(0.06)	-0.01	0.23	0.24(0.13)	-0.02	0.50	0.57(0.32)	-0.04	1.19	-0.38(0.21)	-0.81	0.03	-0.54(0.30)	-1.12	0.04
UNEMP	0.23(0.12)	-0.01	0.46	0.50(0.26)	-0.02	1.02	1.17(0.61)	-0.04	2.36	-0.83(0.45)	-1.74	0.02	-1.07(0.55)	-2.11	0.03
RETD	0.84(0.11)	0.63	1.05	1.83(0.23)	1.39	2.29	4.06(0.48)	3.11	5.01	-3.31(0.44)	-4.19	-2.46	-3.42(0.39)	-4.16	-2.66
NOTINLAB	0.90(0.09)	0.74	1.08	1.97(0.19)	1.61	2.34	4.34(0.39)	3.57	5.11	-3.58(0.36)	-4.29	-2.90	-3.63(0.31)	-4.24	-3.02
HOUSEOWN	-0.52(0.06)	-0.64	-0.40	-1.14(0.14)	-1.40	-0.88	-2.54(0.30)	-3.12	-1.97	2.04(0.25)	1.55	2.54	2.16(0.24)	1.69	2.63

Notes: (*a*) The marginal effect of a dummy variable is estimated as the difference between the probabilities when the dummy is turned on and turned off, keeping all the other variables at their sample mean values. All numbers in this table are presented as percentages. For example the posterior mean for the marginal effect of being male on the probability of poor health status is 0.19%, the posterior standard deviation is 0.09%, and the 95% credibility interval is from 0.04% to 0.38%.

(*b*): This value is -0.0048043243

Table 5
Posterior summary statistics of efficiency for three types of individuals

	r		***E*(*r*)**	
	Mean	**ST.D**	**Mean**	**ST.D**
Best	77.75%	17.75%	77.75%	0.42%
Average	67.46%	23.21%	67.46%	0.31%
Reference	47.65%	29.29%	47.65%	1.25%
Worst	42.69%	29.87%	42.69%	1.34%

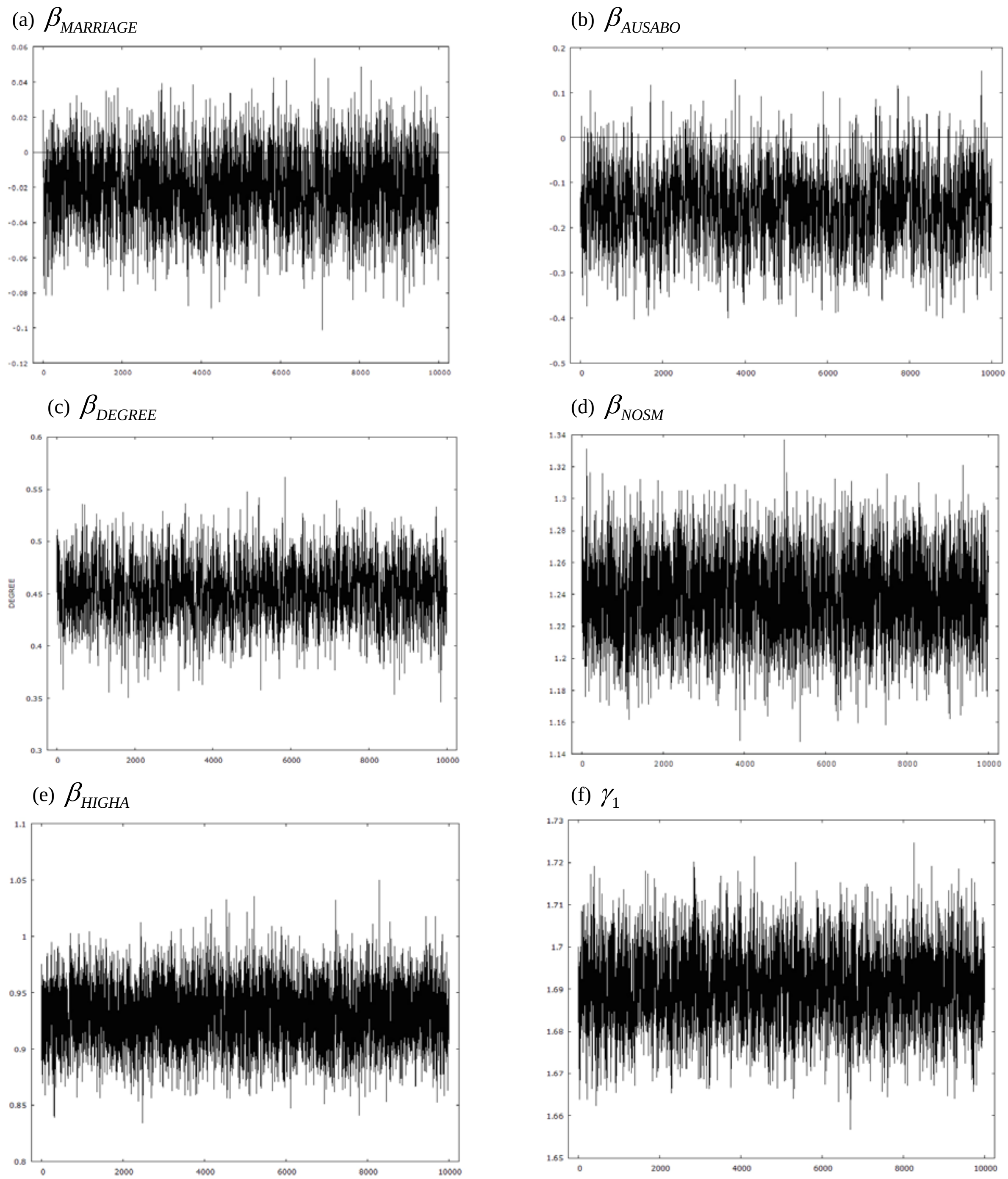


Fig. 1. Sampled path for selected parameters

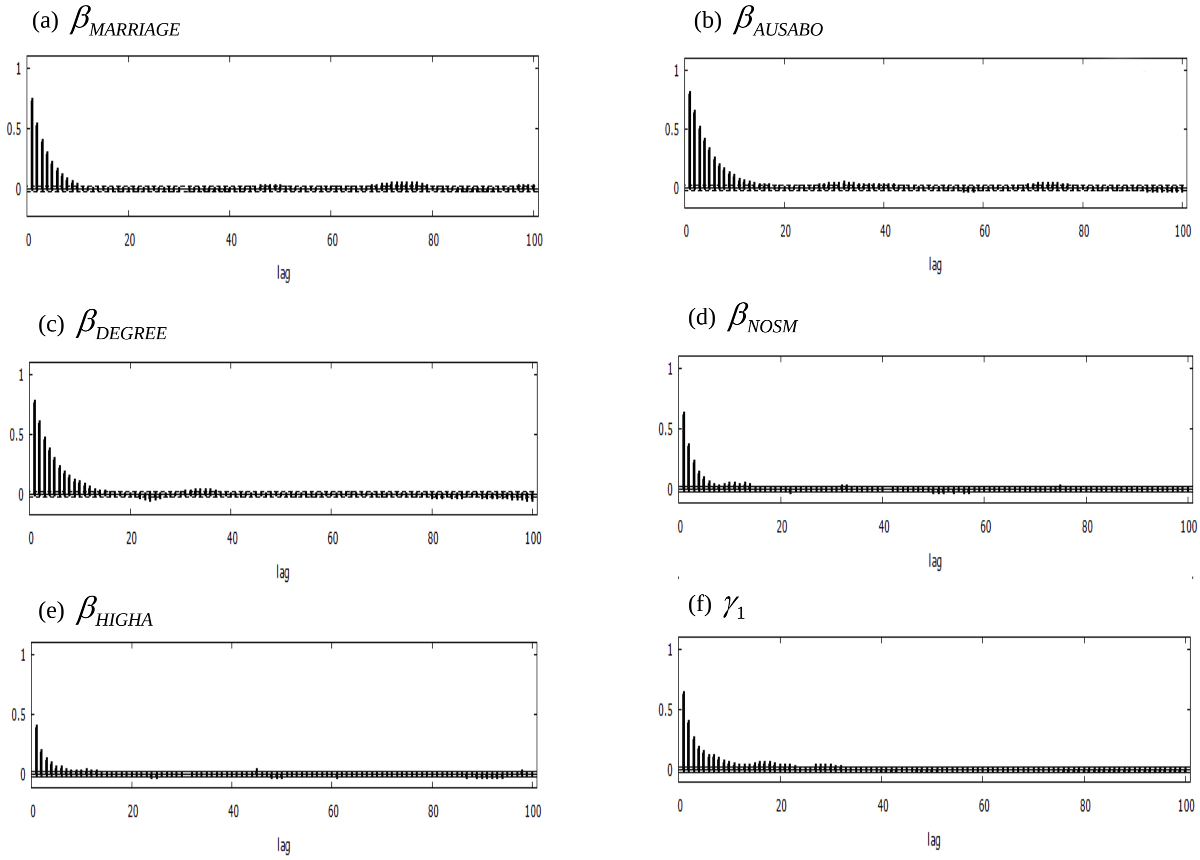
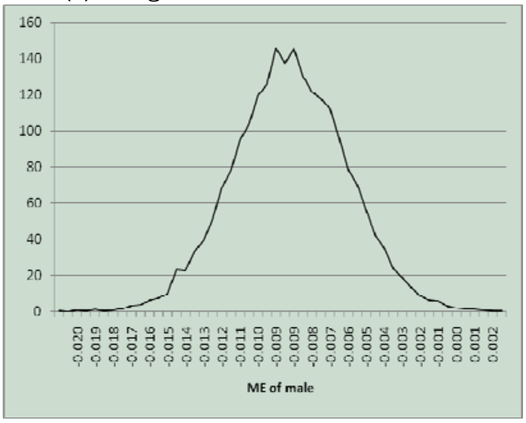
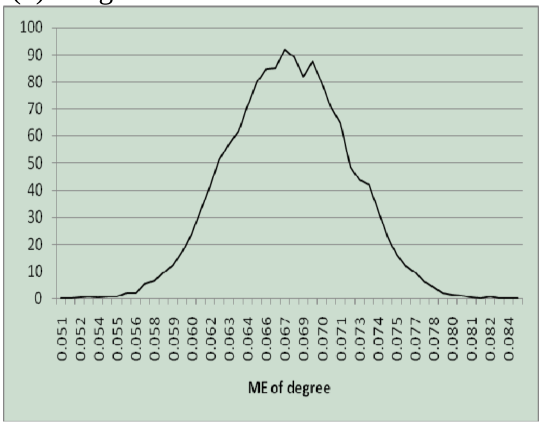


Fig. 2. ACFs for selected parameters

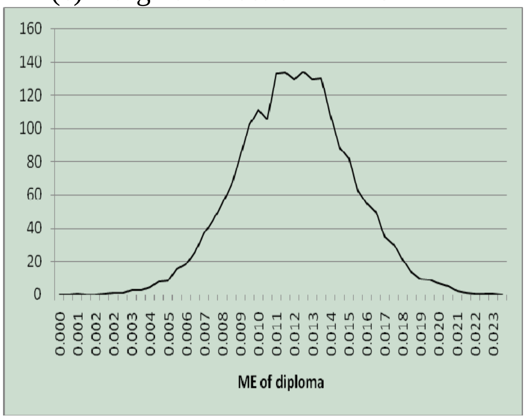
(a) marginal effect of MALE



(b) marginal effect of DEGREE



(c) marginal effect of DIPLOMA



(d) marginal effect of YEAR12

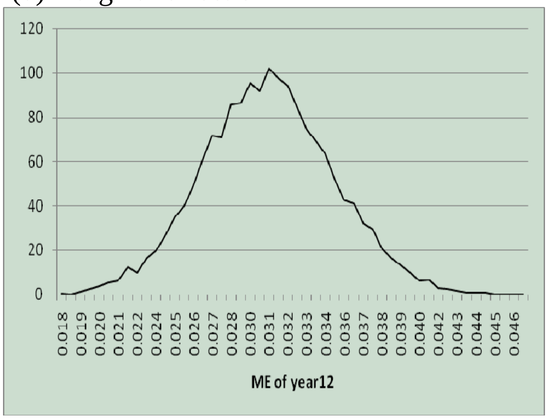
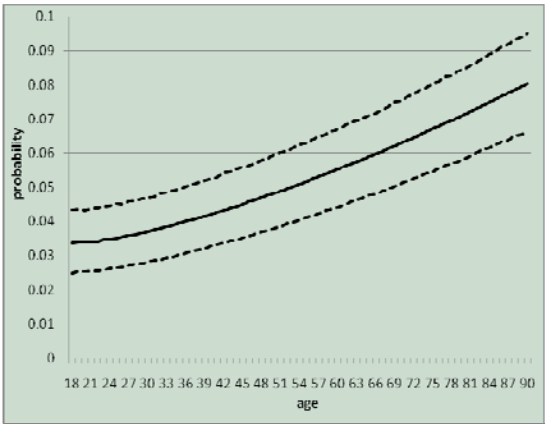
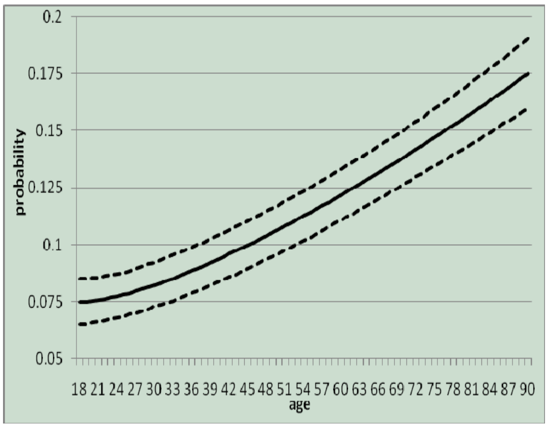


Fig. 3. Posterior densities for marginal effects of selected covariates on the probability of excellent health.

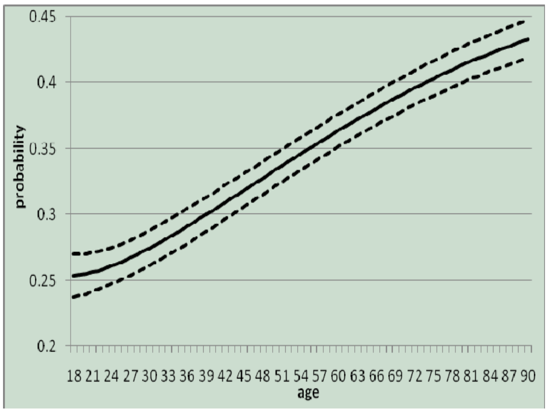
(a) Probabilities of $y = 0$



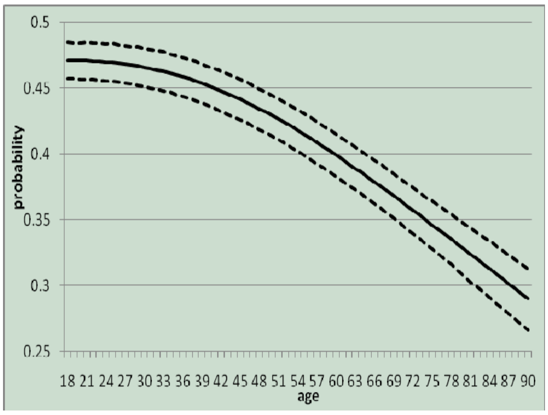
(b) Probabilities of $y = 1$



(c) Probabilities of $y = 2$



(d) Probabilities of $y = 3$



(e) Probabilities of $y = 4$

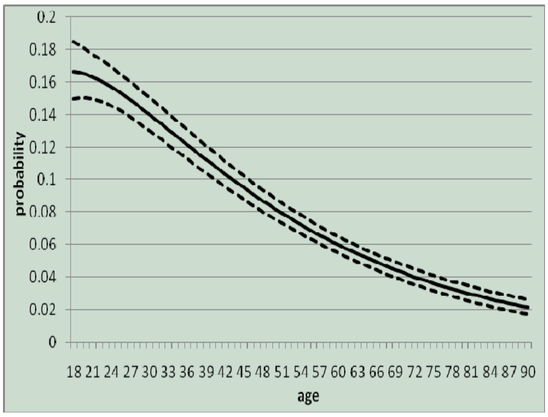


Fig. 4. Probabilities of health status on age. The middle solid line represents the mean of the posterior for the probabilities, the upper dotted line represents the upper 2.5% value of the probability and the lower dotted line represents the lower 2.5% value of the probability.

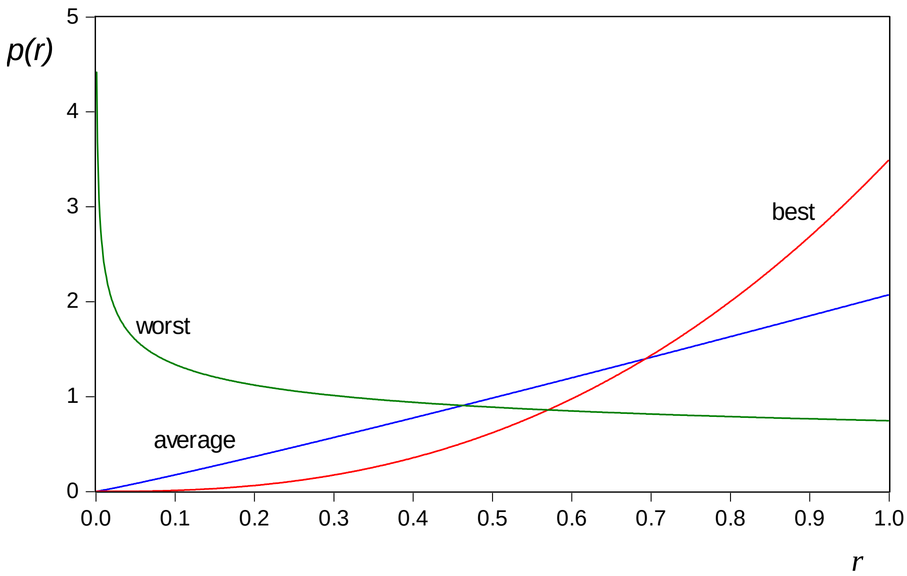


Fig. 5. Predictive densities for efficiency of best, average, and worst settings for *w*.

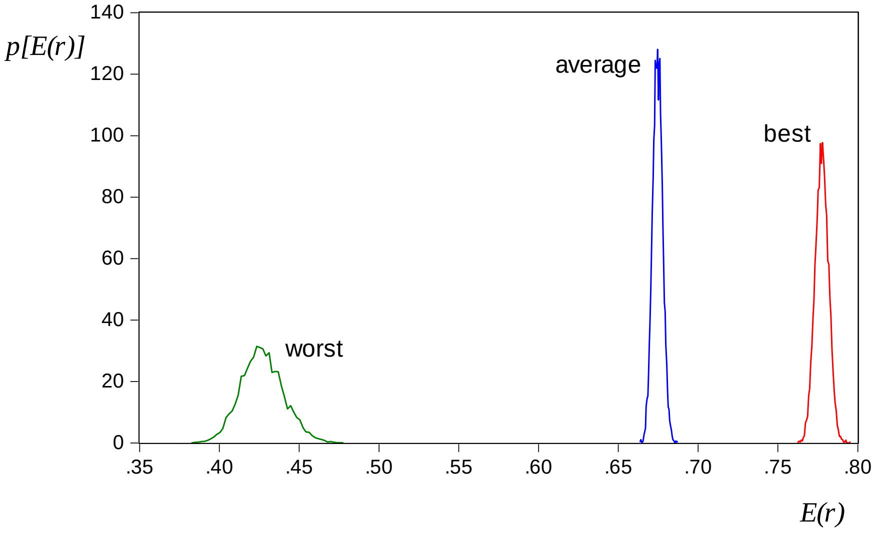


Fig. 6. Posterior densities for mean efficiency of best, average, and worst settings for *w*.

Appendix A. Definition of variables

Variables	Definition
y	
SRH	self-reported health, 0 for poor, 1 for fair, 2 for good, 3 for very good and 4 for excellent
X	
GENDER	1 for male and 0 for female
MARRIAGE	1 if living with somebody in a relationship and 0 otherwise
LNAGE	natural logarithm of age
LNAGE2	square of LNAGE
AUSABO	1 if born in Australia and aboriginal and 0 otherwise
MAINENG	1 if born in other main English speaking countries and 0 otherwise
OTHERC	1 if born in other countries rather than Australia and main English speaking countries and 0 otherwise
AUSNABO	1 if born in Australia and not aboriginal and 0 otherwise. This variable is used as the base for country born status and is dropped in the estimation
DEGREE	1 if the highest qualification is a tertiary degree and 0 otherwise
DIPLOMA	1 if the highest qualification is diploma or trade certificate and 0 otherwise
YEAR12	1 if the highest qualification is Year 12 and 0 otherwise
LOWER12	1 if still in school or cannot finish Year 12 and 0 otherwise. This variable is used as the base for education level and is dropped in the estimation
LONGC1	1 if having long term condition for more than 1 year and 0 otherwise
INNER	1 if living in inner region of Australia and 0 otherwise
OUTER	1 if living in outer region of Australia and 0 otherwise
REMOTE	1 if living in remote region of Australia and 0 otherwise
MAJOR	1 if living in major cities of Australia and 0 otherwise. This variable is used as the base for living region and is dropped in the estimation
STUDENT	1 if full time study and 0 otherwise
PARTTIME	1 if part-time employed and 0 otherwise
UNEMP	1 if unemployed and 0 otherwise
RETD	1 if completely retired from labour market and 0 otherwise
NOTINLAB	1 if not in labour force and 0 otherwise
FULLTIME	1 if part-time employed and 0 otherwise. This variable is used as the base for major activity and is dropped off in the estimation
HOUSEOWN	1 if own or in mortgage of the house and 0 otherwise
W	
NOEX	1 if for all time periods low exercise level [a] and 0 otherwise. This variable is used as the base for exercise and is dropped in the estimation
DOEX	1 if not the case of NOEX and ALDOEX and 0 otherwise
ALDOEX	1 if more than one third of all the time periods high exercise level and less than one third of all the time periods low exercise level
NOSM	1 if never smoke and 0 otherwise
NOA	1 if for all time periods no alcohol risk [b] or low alcohol risk and 0 otherwise
MEDA	1 if not the case of NOA and HIGHA and 0 otherwise. This variable is used as the base for alcohol consumption and is dropped in the estimation
HIGHA	1 if more than one third of all the time periods high alcohol risk and less than one third of all the time periods no alcohol risk or low alcohol risk
LONELY0	1 if never feel lonely [c] and 0 otherwise
LONELY1	1 if sometimes feel lonely and 0 otherwise
LONELY2	1 if always feel lonely and 0 otherwise. This variable is used as the base for social net work and is dropped in the estimation

Notes: (a) We define low exercise level as never doing exercise at all, middle exercise level as doing exercise less than 3 times a week, and high exercise level as doing exercise more than 3 times a week.

(b) Generally, no alcohol risk means 0 standard drinks per week; low alcohol risk means, for males, 1-6 standard drinks per week, or, for females, 1-4 standard drinks per week; high alcohol risk means, for males, at least 7 standard drinks per week, for females, at least 5 standard drinks per week.

(c) The information on loneliness is collected through the statement, ‘I often feel very lonely’. Respondents are assigned a number from 1 to 7 representing from strongly disagree to strongly agree; that is, the higher the number the individual chooses, the more she or he agrees with the statement. We re-classify the respondents into three groups as never feel lonely (1 or 2), sometimes feel lonely (from 3 to 5) and always feel lonely (6 or 7). If this indicator changes over time, we take the value in the last time period.

Appendix B. Percentage of *X* variables by health status

	POOR	FAIR	GOOD	VERY GOOD	EXCELLENT
gender					
FEMALE	3.18	13.98	34.17	36.08	12.59
MALE	3.67	13.96	34.74	35.06	12.57
marital status					
MARRIAGE	3.10	12.81	35.16	36.74	12.19
otherwise	4.07	16.38	32.95	33.22	13.38
living region					
MAJOR	3.13	13.01	33.91	36.50	13.44
INNER	3.77	14.76	34.83	35.14	11.50
OUTER	4.35	17.52	36.07	31.64	10.42
REMOTE	1.89	12.44	35.72	37.23	12.72
health stock					
LONGC1	12.44	34.69	35.13	15.22	2.51
otherwise	0.99	8.40	34.25	41.08	15.29
country born					
AUSABO	7.12	18.03	35.36	28.47	11.02
AUSNABO	3.11	13.81	34.13	36.63	12.32
MAINENG	3.52	12.88	35.10	34.88	13.62
OTHERC	4.71	15.47	35.65	30.75	13.43
education					
DEGREE	1.66	8.22	28.98	42.44	18.70
DIPLOMA	3.32	12.96	36.39	36.25	11.08
YEAR12	2.01	10.55	31.76	39.93	15.76
LOWER12	5.13	19.73	37.15	29.12	8.87
major activity					
STUDENT	1.06	7.71	26.06	42.08	23.09
FULLTIME	0.90	8.37	34.68	40.93	15.13
PARTTIME	1.56	9.73	34.53	40.40	13.79
UNEMP	2.32	17.52	36.56	31.06	12.54
RETD	7.74	27.37	37.28	22.65	4.95
NOTINLAB	9.55	22.50	32.45	26.84	8.65
HOUSEOWN	3.01	13.42	34.51	36.50	12.56
otherwise	4.48	15.45	34.24	33.20	12.63
Overall	3.41	13.97	34.44	35.60	12.58

Appendix C. Descriptive statistics for variables in *W*

	Mean	Std. Deviation
NOEX	0.0511	0.2203
DOEX	0.5637	0.4959
ALDOEX	0.3852	0.4867
NOSM	0.4401	0.4964
NOA	0.1108	0.3139
MEDA	0.7462	0.4352
HIGHA	0.1430	0.3501
LONELY0	0.5737	0.4946
LONELY1	0.3068	0.4612
LONELY2	0.1195	0.3244

References

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

Akin, J. I., 1992. A child health production function estimated from longitudinal data. Journal of Development Economics 38, 323-351.

Albert, J. H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669-679.

Albert, J.H., Chib, S., 1998. Sequential ordinal modeling with applications to survival data. Technical Report. Olin School of Business, Washington University.

Atkinson, S. E., Crocker, T.D., 1992. Econometric health production functions: relative bian from ommited variables and measurement error. Journal of Environmental Economics and Management 22, 12-24.

Baron-Epel, O., Kaplan, G., Haviv-Messika, A., Tarabeia, J., Green, M.S., Kalusli, D.N., 2005. Self-reported health as a cultural health determinant in Arab and Jewish Israelis MABAT - National health and nutrition survey 1999-2001. Social Science & Medicine 61, 1256-1266.

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

Battese, G.E., Coelli, T.J., 1995. A model for technical efficiency effects in a stochastic frontier production function with panel data. Empirical Economics 20, 325-332.

Bauer, P.W., 1990. Recent developments in the econometric estimation of frontiers. Journal of Econometrics 46, 39-56.

Benjamins, M.R., Hummer, R.A., Eberstein, I.W., Nam, C.B., 2004. Self-reported health and adult mortality risk: an analysis of cause-specific mortality. Social Science & Medicine 59, 1297-1306.

Brown, H.S., 2003. Managed care and technical efficiency. Health Economics 12**,** 149-158.

Chen, M-H., Dey, D.K., 1996. Bayesian analysis and computation for correlated ordinal data models. Technical Report #96-42. Department of Statistics. University of Connecticut.

Chen, M-H., Shao, Q-M., Ibrahim, J.G., 2000. Monte Carlo methods in Bayesian computation. Springer-Verlag, New York.

Coelli, T., O'Donnell, C.J., Battese, G.E., Prasada Rao, D.S., 2005. An introduction to efficiency and productivity analysis, Springer, New York, second edition.

Contoyannis, P., Jones, A.M., 2004. Socio-economic status, health and lifestyle. Journal of Health Economics 23**,** 965-995.

Cowles, M.K., 1996. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. Statistics and Computing 6, 101-111.

Desai, S., 1987. The estimation of the health production function for low-income working men. Medical Care 25, 604-615.

Evan, D., Tandon, A., Murray, C., Lauer, J., 2000a. The comparative efficiency of national health systems in producing health : an analysis of 191 countries. World Health Organization, GPE discussion paper, No. 29, EIP/GPE/EQC.

Evan, D., Tandon, A., Murray, C., Lauer, J., 2000b. Measuring overall health system performance for 191 countries. World Health Organization, GPE discussion paper, No. 30, EIP/GPE/EQC.

Fayissa, B., Gutema, P., 2005. Estimating a health production function for Sub-Saharan Africa (SSA). Applied Economics 37, 155-164.

Fernández, C., Osiewalski, J., Steel, M.F.J., 1997. On the use of panel data in stochastic frontier models with improper priors. Journal of Econometrics 79, 169-173.

Gerdtham, U. G., Lothgren, M., Tambour, M., Rehnberg, C., 1999. Internal markets and health care efficiency: a multiple-output stochastic frontier analysis. Health Economics 8, 151-164.

Greene, W., 1990. A gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141-163.

Greene, W., 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Health Economics 13, 959-980.

Greene, W., 2005. Fixed and random effects in Stochastic frontier models. Journal of Productivity Analysis 23, 7-32.

Grossman, M., 1972. Concept of health capital and demand for health. Journal of Political Economy 80, 223-255.

Hakinen, U., Jarvelin, M.R., Rosenqvist, G., Laitinen, J., 2006. Health, schooling and lifestyle among young adults in Finland. Health Economics 15, 1201-1216.

Hollingsworth, B., Wildman, J., 2003. The efficiency of health production: re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. Health Economics 12, 493-504.

Jacobs, P., Schopflocher, D., Klarenbach, S., Golmohammadi, K., Ohinmaa, A., 2004. A health production function for persons with back problems - results from the Canadian Community Health Survey of 2000. Spine 29, 2304-2308.

Jylha, M., 1994. Self-rated health revisited – exploring survey interview episodes with elderly respondents. Social Science & Medicine 39, 983-990.

Kim, S., Shephard, N. and Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65**,** 361-393.

Kim, Y., Schmidt, P., 2000. A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. Journal of Productivity Analysis 14**,** 91-118.

Koop, G., Osiewalski, J. and Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: Hospital cost frontiers. Journal of Econometrics 76, 77-105.

Koop, G., Steel, M.F.J., 2001. Bayesian analysis of stochastic frontier models. In: Baltagi, B.H. (Eds.), A Companion to Theoretical Econometrics. Wiley, Chichester.

Kumbhakar, S.C., Lovell, C.A.K, 2000. Stochastic frontier analysis, Cambridge University Press, New York.

Kumbhakar, S.C., Ghosh, S., McGuckin, J.T., 1991. A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. Journal of Business & Economic Statistics 9, 279-86.

Li, M., Tobias, J. L., 2006. Bayesian analysis of structural effects in an ordered equation system. Studies in Nonlinear Dynamics and Econometrics 10, Article 7.

Liang, J., 1986. Self-reported physical health among aged adults. Journals of Gerontology 41, 248-260.

Meeusen, W., van Den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18**,** 435-444.

Mossey, J. M., Shapiro, E., 1982. Self-rated health - a predictor of mortality among the elderly. American Journal of Public Health 72, 800-808.

Muth, R., 1966. Household production and consumer demand functions. Econometrica 34, 699-708.

Nandram, B., Chen, M-H., 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. Journal of Statistical Computation and Simulation 54, 129-144.

Puig-Junoy, J., 1998. Measuring health production performance in the OECD. Applied Economics Letters 5, 255-259.

Puig-Junoy, J., Ortun, V., 2004. Cost efficiency in primary care contracting: a stochastic frontier cost function approach. Health Economics 13**,** 1149-1165.

Rosko, M.D., 2001. Cost efficiency of US hospitals: a stochastic frontier approach. Health Economics 10, 539-551.

Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57-66.

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

Thornton, J., 2002. Estimating a health production function for the US: some new evidence. Applied Economics 34, 59-62.

Van den Brink, C.L., Tijhuis, M., van den Bos, G.A.M., Giampaoli, S., Nissinen, A., Kromhout, D., 2005. The contribution of self-rated health and depressive symptoms to disability severity as a predictor of 10-year mortality in European elderly men. American Journal of Public Health 95, 2029-2034.

van den Broeck, J., Koop, G., Osiewalski, J., Steel, M.F.J., 1994. Stochastic frontier models – a Bayesian perspective. Journal of Econometrics 61, 273-303.