American Journal of Agricultural Economics Advance Access published November 21, 2014

A Birth-Death Markov Chain Monte Carlo Method to Estimate the Number of States in a State-Contingent Production Frontier Model

José Luis Preciado Arreola and Andrew L. Johnson

In this article we estimate a state-contingent production frontier for a group of farms while endogenously estimating the number of states of nature induced by unobserved environmental variables. This estimation is conducted by using a birth-death Markov chain Monte Carlo method. State-contingent output is estimated conditioned on an observed input vector and an a priori unknown number of unobserved states, each of which is modeled as a component of a mixture of Gaussian distributions. In a panel data application, state-independent dummy variables are used to control for time effects. The model is applied to 44 rice farms in the Philippines operating between 1990 and 1997. The endogenous estimation procedure indicates a unimodal posterior probability distribution on the number of states, with a median of three states. The estimated posterior coefficient values and their economic implications are compared to those of previous research that had assumed a fixed number of states determined exogenously. Goodness-of-fit testing is performed for the first time for a state-contingent production model. The results indicate satisfactory fit and also provide insights regarding the degree of estimation error reduction achieved by utilizing a distribution for the number of states instead of a point estimate. All of our models show significant improvement in terms of mean squared error of in-sample predictions against previous work. This application also demonstrates that using a state-independent dummy time trend instead of a state-contingent linear time trend leads to slightly smaller differences in state mean output levels, although input elasticities remain state-contingent.

Key words: Bayesian SFA, birth-death Monte Carlo, state-contingent production, stochastic frontier analysis.

JEL codes: C3, C5, C6, Q1.

State-contingent production provides a flexible framework for estimating production functions (Quiggin and Chambers 2006), particularly in industries with subsets of firms that operate in environments lacking data at the producer level. Using rice farming as an example, data on each farm's rainfall, temperature, pests, diseases, and other environmental factors are potentially observable, but rarely collected at the farm level. It is possible, however, to employ state-contingent production frontiers to separate the impacts of inefficiency and state-dependent production conditions by allowing different parameters for each state of nature. The simultaneous estimation of production frontiers using stochastic frontier analysis (SFA), for example, and the creation of unobserved variable clusters that are needed to define the states of nature can be addressed by using latent class stochastic frontier models (LCSFM), either from a sampling framework—an example of which appears in Orea and Kumbhakar (2004)—or from a Bayesian framework via Markov chain Monte Carlo (MCMC) methods. We select the Bayesian framework because it allows us to obtain the complete posterior distribution of the number of states and to calculate weights for our output predictions. Moreover, MCMC methods make it relatively easy to add common constraints, such as monotonicity in inputs, to the production

José Luis Preciado Arreola is a PhD student and Andrew L. Johnson is an associate professor, both in the Department of Industrial and Systems Engineering, Texas A&M University. Correspondence may be sent to: ajohnson@tamu.edu.

The authors would like to thank the seminar audiences at Aalto University and the North American Productivity Workshop. José Luis Preciado Arreola is partially supported by the Mexican National Council of Science and Technology (CONACYT), scholarship number 209414.

Amer. J. Agr. Econ. 1–19; doi: 10.1093/ajae/aau096

frontier estimation algorithm, and to model the observed outputs as the outcome of a finite mixture of Gaussian distributions.

Several uses of the state-contingent approach have been described in the production economics literature. For example, Chavas (2008) develops an input cost-minimizing methodology under a state-contingent technology. Further, Nauges, O'Donnell, and Quiggin (2011) use the state-contingent approach and MCMC methods to model production with Constant Elasticity of Substitution (CES) flexible production functions. Other studies include O'Donnell and Shankar (2010) and Serra and Lansink (2010). O'Donnell and Griffiths (2006) estimate a state-contingent frontier model using MCMC methods, specifically a Gibbs sampler, for the same data set we analyze. O'Donnell and Griffiths determine the number of states of nature by minimizing the Bayesian Information criterion (BIC) for a model in which only the intercepts of the production function and the precision parameters (the inverses of the variances of the noise distribution) are state-contingent. These authors then impose the BIC-minimizing number of states to a model for which the input slopes are also state-contingent and monotonicity of the production function is imposed. We caution, however, that the BIC-minimizing number of states can differ across models since the number of parameters increases for models in which the state-contingent input slopes are estimated. In other words, the number of states imposed may be inappropriate even though imposing the number of states calculated for a simpler model is computationally easier.

Further, in the frontier-estimation models discussed above, the number of states is either fixed or obtained by using BIC as model selection criterion. Biernacki, Celeux, and Govaert (1998) observe that the assumptions needed to use BIC as a model selection criterion do not hold for mixture models. For example, the assumption that all model parameters are well inside of the parameter space is violated when evaluating models for which the number of states is larger than the true one, as state probabilities for some states will approach zero, which is a boundary value for a probability parameter, making BIC inappropriate. Biernacki, Celeux, and Govaert (2000) and Biernacki and Govaert (1997) show that other criteria such as the Integrated Classification Likelihood (ICL) and the Classification Likelihood Criterion (CLC) are often better-suited for this type of application. Banfield and Raftery (1993) show that approximate weight of evidence (AWE), derived as a Bayes Factor approximation, can be used to compute the posterior probabilities for each number of states given the data. It is also possible to use these posterior probabilities to weight the chosen model instead of simply choosing one with the best score. Regardless of the model selection criterion chosen, a two-stage procedure is required, that is, the first stage estimates all plausible models, and the second stage scores and obtains the posterior probabilities to weight the models.

In contrast, endogenous estimation of the number of states involves solving the model selection and estimation problems simultaneously. Specifically, the selection infers the number of states in the model and estimation calculates the coefficients for the regression models given a certain number of states. Both endogenous estimation and the model selection criteria mentioned above allow us to identify the best model; however, endogenous estimation also calculates weighted output across all plausible models in a single run of the estimation algorithm. Moreover, though in both the endogenous estimation and the two-stage procedure described above the range of possible values for the number of states has to be specified a priori, in the endogenous estimation algorithm presented in this article, only the plausible values of this range are visited by the estimation algorithm, thus significantly reducing computational complexity.

Our survey of the literature indicates that efforts to endogenously estimate the number of components in a Gaussian mixture in an econometric context have been limited. For example, in a macroeconomic setting Chopin and Pelgrin (2004) developed a general hidden Markov chain and estimated all transition probabilities where the number of transition probabilities grows quadratically in the number of states. Other methods can be used to estimate a mixture model where the number of components is unknown, such as the reversible jump Markov chain Monte Carlo (RJMCMC) introduced by Green (1995) and first applied to mixture modeling by Richardson and Green (1997). Nevertheless, Richardson and Greene's RJMCMC algorithm is not directly applicable in a regression context, because the equations for the parameter means become datapoint-specific and cause estimation difficulties (Kottas and Gelfland 2001). Some efforts to tailor RJMCMC algorithms within a regression context have been successful, such as Denison, Mallick, and Smith (1998) and Biller (2000).

However, we select the birth-death Markov chain Monte Carlo (BDMCMC) method (Stephens 2000), which was first used in a regression context by Hurn, Justel, and Robert (2003). The BDMCMC computes the rates of a simpler hidden Markov Chain, which these authors characterize as a birth and death process. By reducing the number of parameter estimates, the use of BDMCMC is better-suited for applications with limited data. We also note that the computational process becomes simpler compared to RJMCMC because the use of BDMCMC reduces the types of jumps across states while still achieving consistent estimation results.

To the best of our knowledge, this is the first study to endogenously estimate the number of states in a state-contingent stochastic production frontier, as well as the first to perform a goodness-of-fit analysis on a state-contingent production frontier model. We apply the proposed BDMCMC method to a case study of 44 rice farms in the Philippines operating between 1990 and 1997, which were investigated in Villano, O'Donnell, and Battese (2004) and later in O'Donnell and Griffiths (2006). In addition, we relax the linear time trend assumption of the models in O'Donnell and Griffiths for changes in production over time to consider a model with a state-independent dummy time trend, which more accurately separates the frontier shifts over time from state-contingent yield variability, inefficiency, and noise. We discuss the implications of using both the linear time trend and dummy time trend models on the estimated posterior parameters, the relevant economic interpretations, and the existence of state-specific production. We describe the use of mean squared error (MSE) and visual tools, such as quantile-quantile (QQ) plots, to compare the performance of a model using the full distribution of the number of states versus a model using only the mode of this distribution.

The remainder of this article is organized as follows. The model section presents the functional form and our assumptions about the production frontier, the dependence and distributional assumptions about the frontier parameters to be estimated, and the steps followed to conduct inference. The empirical application section describes the rice farm case study, including the application-specific prior parameter values, the estimation results, our economic interpretations, and our incorporation of some common economic constraints. We then test the fit and robustness of our results and discuss our sensitivity analyses. We also assess the degree of label-switching present in our results. The conclusions section, which summarizes our findings, mentions some possible future research pursuits.

Model

We estimate a production frontier and the number of states using a BDMCMC algorithm with a "nested" Gibbs sampler. This Gibbs sampler draws from the conditional posterior distributions of the parameters to be estimated and infers the frontier coefficients, precisions, state probabilities, state allocations, and inefficiency terms of the observations in the dataset for a fixed number of states. To obtain expressions that are proportional to the conditional posterior distributions of each parameter, we must multiply the joint prior distribution against the likelihood function—the joint density of all output levels in our dataset conditioned on all other parameters—to obtain an unnormalized joint posterior distribution. The joint prior distribution is defined by the assumed prior distributions on all parameters and by the Bayesian Hierarchical model that specifies the dependence structure between them.

The BDMCMC algorithm selects the number of states by computing the birth and death rates for states through likelihood comparisons (see step 1 of the algorithm in the Bayesian Inference via BDMCMC subsection below) between the current model and models where an individual state is removed. Once these rates are obtained at each pass of the BDMCMC algorithm, a birth or death happens within a Markov chain describing the number of states. Stephens (2000) proves that the stationary distribution of this Markov chain, obtained after a large enough number of births and deaths, converges to the correct posterior distribution for the number of states. Moreover, Stephens (2000) proves that using the Gibbs sampler to obtain draws from all frontier parameters each time a state is born or dies results in a valid joint distribution of all of the parameters estimated within the BDMCMC algorithm, which simultaneously solves both the estimation and model selection problems.

Production Frontier Regression Model

We first define a state-contingent production frontier regression model (the production frontier) that will be estimated by our BDMCMC algorithm

$$(1) \qquad \ln Y = f_j(X_1, \ldots, X_k) + \varepsilon$$

where f_j is the production function associated with state j, and $\varepsilon = v - u$ is a composite error term obtained by subtracting a nonnegative variable u representing the technical inefficiency of a farm relative to the efficient frontier from a random effect, v, which follows a finite normal scale mixture distribution centered at 0. We assume that all f_j's have the same functional form; however, the production frontier coefficients are state-dependent following O'Donnell and Griffiths' most general random effects model. We scale all input variables to their mean values and adapt the model for a panel context, resulting in

$$(2) \qquad \ln Y_{it} = \boldsymbol{d}_{it}\boldsymbol{\beta}_{\mathbf{0}} + (\boldsymbol{x}_{it} \otimes \boldsymbol{d}_{it})'\boldsymbol{\beta}_{\mathbf{-0}} + v_{it} - u_i.$$

Note that we assume a fully separable intercept term $\boldsymbol{\beta}_{\mathbf{0}}$ from the chosen functional form. To have state-varying coefficients, we define J to be the number of states, and $\boldsymbol{d}_{it}$ as a vector containing J state-specific binary dummy variables. Moreover, $\boldsymbol{\beta}_0 = [\beta_{01}, \ldots, \beta_{0J}]'$ is a vector of state-varying intercepts, $\otimes$ is the Kronecker product, $\boldsymbol{\beta}_{-0} = [\boldsymbol{\beta}_{\mathbf{-01}}, \ldots, \boldsymbol{\beta}_{\mathbf{-0J}}]'$ is a vector of state-varying slope coefficients, and $\boldsymbol{\beta} = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_{\mathbf{-0}}]'$ is a vector containing the coefficients of all J frontiers. The $\boldsymbol{d}_{it}$ vector is used to specify the distributions of the random effects, $v_{it} \sim N(0, [\boldsymbol{d}_{it}\boldsymbol{h}]^{-1})$, where $\boldsymbol{h}$ is the vector of state-specific precisions. Finally, the output vector $\boldsymbol{Y_{it}}$ is the rice yield of farm i at time t.

Next, we consider two distinct time trend assumptions and fit separate models for each: in the first model, the linear time trend model, β_{TRj} (included in $\boldsymbol{\beta}_{\mathbf{-0j}}$) is a scalar and in the second model (the dummy time trend model) it is a vector of state-independent dummy coefficients for each year, that is, $\boldsymbol{\beta}_{TRj} = [\beta_{TR2j}, \ldots, \beta_{TRTj}]$, with $\beta_{TRtj} = \beta_{TRtk} \forall j, k \in \{1, \ldots, J\}$, for each year $t \in \{2, \ldots, T\}$ in the time span. The dummy time trend model is important because the results produced by models with less flexible time trends, that is, the linear trend model and the models estimated by O'Donnell and Griffiths, could lead us to attribute output variations over time to the state-specific effects on mean output, to noise, and/or to inefficiency.

Bayesian Hierarchical Model

To estimate the parameters of the production frontier with a Gibbs sampler, as done in our application of BDMCMC, we define a dependence structure using a Bayesian Hierarchical model. Figure 1 illustrates this hierarchical structure; the circles indicate parameters to be inferred during the BDMCMC algorithm, the squares indicate known or assumed quantities, and the arrows indicate that the value of the parameter located at the end of an arrow depends on the value of the parameter at its beginning (see the prior and conditional posterior distributions subsections for details of the explicit mathematical prior and posterior dependence among the parameters).

Similar to Richardson and Green (1997), our hierarchical model extends the models in O'Donnell and Griffiths by including g and m as hyperparameters of Θ, which allows for smooth precision estimates among the states. Similar to O'Donnell and Griffiths and

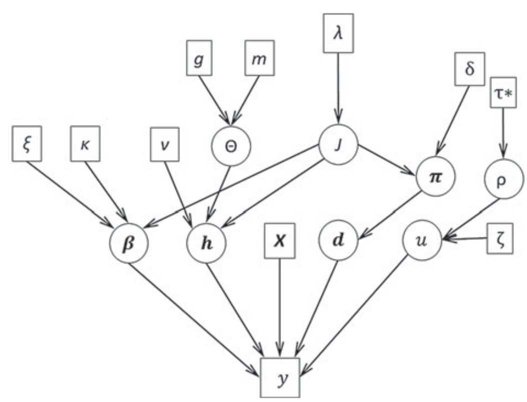

Figure 1. Directed acyclic graph for the Bayesian hierarchical model used for parameter estimation, with parameters to be inferred in circles

Hurn, Justel, and Robert (2003), the output is dependent on the input matrix $\boldsymbol{X}$, leading to a standard regression structure. For clarity, we include J, the number of states in the model, and the hyperparameter for its prior mean, λ. Further, we define $\boldsymbol{\pi}$, a vector of state probabilities (or the probability that any given observation is assigned to a particular component of the mixture), and ρ to be a hyperparameter of u related to median efficiency. Finally, note that the dependence structure detailed in figure 1 can also be described as the joint prior distribution, given by

$$
\begin{aligned}
(3)\qquad & p(\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{d}, u, \boldsymbol{\pi}, \rho, \Theta, J) \\
&= p(\boldsymbol{\beta}|J, \xi, \kappa) \cdot p(\boldsymbol{h}|J, v, \Theta) \cdot p(\boldsymbol{\pi}|J, \delta) \\
&\quad \cdot p(\boldsymbol{d}|\pi) \cdot p(u|\rho, \zeta) \\
&\quad \cdot p(\Theta|g, m) \cdot p(\rho|\tau^*) \cdot p(J|\lambda)
\end{aligned}
$$

and the likelihood function $p(y|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{d}, u, \boldsymbol{\pi}, \rho, \Theta, J)$ Note that output, y, appears in the likelihood function, which is further discussed in the likelihood function section below, but not in the joint prior distribution.

Prior Probability Distributions

To obtain conditional posterior distributions from which the Gibbs sampler will draw, we define the prior probability distributions for the parameters to be estimated. The prior distributions are dependent on the known prior hyperparameter values. Our production frontier includes a large number of parameters to be estimated, and thus requires us to define proper prior distributions on all parameters. A commonly used prior distribution for the regression coefficients and precisions of a production frontier in a Bayesian framework is the normal-inverse gamma (NIG). We also use the prior distributions to impose structure to the states being estimated. Although a loss function has been used to address the state identifiability problem (Hurn, Justel, and Robert 2003), we use a labeling restriction that ranks the states of nature from "least favorable" to "most favorable" in order to gain knowledge about the relative desirability of each state. We define our labeling restriction to constrain the expected log-output from a lower-indexed state to be, on average, less than that of a higher-indexed state with the NIG assumption to obtain $\boldsymbol{\beta} \sim N(\boldsymbol{\xi}, \boldsymbol{\kappa}) \cdot I(E(\ln Y_{it}|x_{it}=0, j=1) \leq \cdots \leq E(\ln Y_{it}|x_{it}=0, j=J))$, where $\boldsymbol{\xi}, \boldsymbol{\kappa}$ are fixed and the indicator function describes the labeling restriction.

Consistent with the NIG prior, we assign a gamma prior for each state-specific precision, h_j, such that $h_j \sim \Gamma(v, \Theta)$ for all $j \in \{1, \ldots, J\}$, where its hyperparameter Θ has a prior $\Gamma(g, m)$ distribution, and v is fixed. We choose a uniform prior for $\boldsymbol{\pi}$ over J states, a Multinomial($\boldsymbol{\pi}$) distribution for $\boldsymbol{d}$, and following Stephens (2000), we choose the number of states to follow a prior Poisson distribution, $P_0(\lambda)$. Also, we assume a (properly normalized) truncated exponential prior distribution for the inefficiency term of the ith farm, u_i. Thus, $u_i \sim c \cdot \text{expo}(\rho) \cdot I(u_i \leq -\log(\zeta))$, where c is a normalization constant and $-\log(\zeta)$ is an upper bound on the inefficiency term. Note that this upper bound translates into a lower bound ζ on the technical efficiency of each farm (TE_i) due to $TE_i = \exp(-u_i)$[1]. Finally, we assume that ρ, the prior hyperparameter of the u_i terms, follows an expo $(-1/\ln(\tau^*))$ distribution, where we assume that τ^* is a prior estimate of the median technical efficiency.

Likelihood Function

Having specified the prior distributions and dependence structure for our parameters, we now need to specify a likelihood function to multiply with them in order to obtain expressions that are proportional to the conditional posterior parameter distributions that our Gibbs sampler will draw from. We construct the standard likelihood for a mixture of Gaussians with unobserved state allocations by weighting the Gaussian likelihoods with different coefficients and precisions, each of which corresponds to a particular component in the mixture, times the component's relative frequency within the mixture. In equation (3) below, the total number of farms is denoted by N, and the index of the last time period is denoted by T. Below, $\boldsymbol{y}$ denotes the vector of all observations $\ln(Y_{it})$. Note that we use "..." to denote conditioning an all parameter models except for the one of interest:

[1] While justifications exist for this in some industries, in this paper this assumption is needed for estimation proposes.

$$(4) \quad y|\ldots \sim (2\pi)^{-\frac{NT}{2}} \prod_{i=1}^{N}\prod_{t=1}^{T} \left\{\sum_{j=1}^{J} \pi_j \sqrt{h_j} \exp\left[-0.5h_j(\ln(Y_{it}) - \beta_{0j} - \boldsymbol{x}'_{it}\boldsymbol{\beta}_{-0j} + u_i)^2\right]\right\}.$$

Conditional Posterior Distributions

We multiply the likelihood against the joint prior distribution assumed for our parameters to obtain an expression that is proportional to their joint posterior distribution, which we then condition. To obtain valid parameter estimates, given a fixed J, for the production frontier coefficients $\boldsymbol{\beta}$, precisions $\boldsymbol{h}$, state probabilities $\boldsymbol{\pi}$, the dummy variables to assign observations to states $\boldsymbol{d}$, the inefficiency terms u_i, and hyperparameters Θ and ρ, we need to draw from their conditional posterior distributions using the Gibbs sampler until the parameter estimates reach stationarity. For our production frontier model, and given our choices of dependence structure, prior distributions and likelihood, the conditional posteriors for $\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\pi}$, and $\boldsymbol{d}$ correspond to the results of a NIG finite Gaussian mixture model for fixed J, while the ones corresponding to u_i, Θ, and ρ are straightforward to derive. Note that the conditional posterior distribution for u_i is not conjugate, it is a 2-sided truncated normal, but has a truncated exponential prior distribution. For simplicity, we define $\boldsymbol{z_{it}} = (\boldsymbol{d}'_{it}, (\boldsymbol{x_{it}} \otimes \boldsymbol{d_{it}})')'$. Noting that parameters n_j refer to the number of observations allocated to state j, the conditional posterior distributions for our parameters are given by the following:[2]

$$(5) \quad \boldsymbol{\pi}\ldots \sim D(\delta + n_1, \ldots, \delta + n_k)$$

$$(6) \quad \boldsymbol{\beta}|\ldots \sim N(\bar{\boldsymbol{\xi}}, \bar{\boldsymbol{\kappa}}) \cdot I(E(\ln Y_{iy}|x_{it} = 0, j = 1) \leq \cdots \leq E(\ln Y_{it}|x_{it} = 0, j = J))$$

$$(7) \quad h_j|\ldots \sim \Gamma\left\{v + \frac{1}{2}n_j,\ \Theta + \frac{1}{2}\sum_{it:z_{it}=j}[y_i + u_i - (\beta_{0j} + \boldsymbol{x_{it}}\boldsymbol{\beta}_{-0j})]^2\right\}$$

$$(8) \quad \boldsymbol{d}|\ldots = \prod_{i=1}^{N}\prod_{t=1}^{T} f_M(\boldsymbol{d_{it}}|1, \bar{\boldsymbol{d}}_{\boldsymbol{it}})$$

$$(9) \quad \Theta|\ldots \sim \Gamma\left(g + J\nu, \mathrm{m} + \sum_j h_j\right)$$

$$(10) \quad u_i|\ldots \sim \mathrm{N}(\mu_{ui}\sigma^2_{ui}) \cdot I(0 \leq u_i \leq -\log(TE_l))$$

and

$$(11) \quad \rho^{-1}|\ldots \sim \Gamma\left(\frac{\mathrm{N}+1}{\boldsymbol{u}'\boldsymbol{j}_N - \ln(\tau^*)}, 2(N+1)\right).$$

Bayesian Inference via BDMCMC

Having specified the conditional posterior distributions from which we will sample, the last step is to simultaneously estimate the parameters of our state-contingent production frontier model and its number of states. Prior to explaining this, we will discuss some technicalities of the random variable simulation process conducted in our Gibbs sampler, as standard random variable generators do not suffice simulating our model's parameters. Efficient sampling from the restricted normal distribution of $\boldsymbol{\beta}$ becomes critically important as the dimensionality of the model increases for larger values of J, therefore we incorporate an efficient sampling method (Geweke 1991) when the standard accept-reject criterion fails to produce a satisfactory draw from $\boldsymbol{\beta}$'s posterior distribution after 30 trials. We select this number to ensure a reasonable running time, due to frequent sampling from the simple accept-reject sampler, and because Geweke's sampler produces draws whose variance depends on the number of iterations of its internal algorithm, which we want to avoid where possible. Unlike O'Donnell and Griffiths, we use the 2-sided truncated normal sampler (Robert 1995) to simulate from the posterior distribution. Finally, we nest our Gibbs sampler within the BDMCMC algorithm and solve the model estimation and model selection problems simultaneously by adapting the algorithm outlined in Stephens (2000). Letting $\boldsymbol{\eta}^{(s)} = (\boldsymbol{\beta}^{(s)}, \boldsymbol{h}^{(s)})'$, the steps in our BDMCMC algorithm are as follows:

1. Run the birth and death process, starting at time s for a fixed time s_0, fixing $\boldsymbol{d}, \rho^{-1}, \boldsymbol{u}, \Theta, \boldsymbol{\pi}$, and $\boldsymbol{\eta}$. Keep $J^{(s+s_0)}$ as a draw from the conditional posterior of J.

    a) Fix a birth rate $\lambda b = \lambda$.

[2] See the online supplementary appendix for the posterior hyperparameter expressions and the data.

b) Compute death rate for each component $\delta_j(\boldsymbol{\pi},\boldsymbol{\eta}) = \frac{L((\boldsymbol{\pi},\boldsymbol{\eta})/(\pi_j,\eta_j))}{L(\boldsymbol{\pi},\boldsymbol{\eta})} \forall j$.

c) Compute total death rate $\delta(\boldsymbol{\pi},\boldsymbol{\eta}) = \sum_j \delta_j(\boldsymbol{\pi},\boldsymbol{\eta})$.

d) Compute next time s_{new} until a birth or death occurs from an $expo(\lambda_b + \delta(\boldsymbol{\pi},\boldsymbol{\eta}))$ distribution, and let $s^* = s + s_{new}$.

e) Decide type of jump: birth with probability $\frac{\lambda_b}{\lambda_b+\delta(\boldsymbol{\pi},\boldsymbol{\eta})}$ or death with probability $\frac{\delta(\boldsymbol{\pi},\boldsymbol{\eta})}{\lambda_b+\delta(\boldsymbol{\pi},\boldsymbol{\eta})}$.

f) Adjust sizes of $\boldsymbol{d}$, $\boldsymbol{\pi}$, and $\boldsymbol{\eta}$ according to the type of jump.
For birth jumps, generate $\boldsymbol{\pi}$ according to a $Beta(1,k)$ distribution and generate a new component for $\boldsymbol{\eta}$ from its prior. Assign the new component an index j^* such that it does not violate the labeling restriction. For death jumps, select a component to die with a probability of $\delta_j(\boldsymbol{\pi},\boldsymbol{\eta})/\delta(\boldsymbol{\pi},\boldsymbol{\eta})$.

g) Repeat until $s^* \geq s + s_0$.

2. Draw $\boldsymbol{d}^{(s+1)}$ from $p(\boldsymbol{d}|J^{(s+1)}, \boldsymbol{\pi}^{(s+s_0)}, \boldsymbol{\eta}^{(s+s_0)}, \Theta^{(s)}, \rho^{-1(s)}, \boldsymbol{u}^{(s)})$.
3. Draw $\Theta^{(s+1)}$ from $p(\Theta|J^{(s+1)}, \boldsymbol{\pi}^{(\boldsymbol{s+s_0})}, \boldsymbol{\eta}^{(s+s_0)}, \boldsymbol{d}^{(s+1)}, \rho^{-1(s)}, \boldsymbol{u}^{(s)})$.
4. Draw $\boldsymbol{\pi}^{(s+1)}$ from $p(\boldsymbol{\pi}|J^{(s+1)}, \boldsymbol{\eta}^{(s+s_0)}, \boldsymbol{d}^{(\boldsymbol{s+1})}, \Theta^{(s+1)}, \rho^{-1(s)}, \boldsymbol{u}^{(s)})$.
5. Draw $\boldsymbol{\eta}^{(s+1)}$ from $p(\boldsymbol{\eta}|J^{(s+1)}, \boldsymbol{\pi}^{(s+1)}, \boldsymbol{d}^{(s+1)}, \Theta^{(s+1)}, \rho^{-1(s)}, \boldsymbol{u}^{(s)})$.
6. Draw $\rho^{(s+1)}$ from $p(\rho|J^{(s+1)}, \boldsymbol{\pi}^{(s+1)}, \boldsymbol{\eta}^{(s+1)}, \boldsymbol{d}^{(s+1)}, \Theta^{(s+1)}, \boldsymbol{u}^{(s)})$.
7. Draw $\boldsymbol{u}^{(s+1)}$ from $p(\boldsymbol{u}|J^{(s+1)}, \boldsymbol{\pi}^{(s+1)}, \boldsymbol{\eta}^{(s+1)}, \boldsymbol{d}^{(\boldsymbol{s+1})}, \Theta^{(s+1)}, \rho^{-1(s+1)})$.

Note that the choice of λ_b is inconsequential for the algorithm to converge because λ_b can take any finite positive value (Stephens 2000). Since the birth or death jumps of the algorithm must consider the labeling restriction, we add another instruction to step 1(f), which assigns the newly-generated index component j^* such that its intercept is larger than that of the $(j^*-1)th$ component and smaller than that of the $(j^*+1)th$ component. Finally, note that drawing from our Gibbs sampler takes place during steps 2-5.

Empirical Application

Our dataset contains observations for 44 rice farms operating in the Tarlac region of the Philippines between 1990 and 1997. Each observation, which refers to a farm i at year t, comprises the value for the yield (in tons), the output variable Y, and the values for area planted (hectares planted), labor used (person-days), and fertilizer used (kilograms of nitrogen, phosphorus, and potassium or NPK fertilizer), the input variables, X_1, X_2, and X_3, respectively. Summary statistics for our database, identical to the ones shown in O'Donnell and Griffiths, are shown in table 1. We use a translog production function, for which the input vector is

$$(12) \quad \boldsymbol{x_{it}} = [TR_{it}, \ln(X_{1it}), \ln(X_{2it}), \ln(X_{3it}), 0.5\ln(X_{1it})^2, \ln(X_{1it})\ln(X_{2it}) \ln(X_{1it})\ln(X_{3it}), 0.5\ln(X_{1it})^2, \ln(X_{2it})\ln(X_{3it}), 0.5\ln(X_{3it})^2]'$$

where TR is either a scalar or a vector of the binary dummy variables, depending on which model we consider. We scale all input variables at their means and define all of the dummy variables of the dummy time trend model as the difference in the yearly shift of the production year against the base year of 1990 following Baltagi and Griffin (1988); this base year has no associated time dummy variable. Since the variables are scaled at their input means, ordering states by their intercept values ensures that the mean log-output of a higher-indexed state will be greater than that of a lower-indexed state, that is, $I(\beta_{01} \leq \beta_{02} \leq \cdots \leq \beta_{0J})$ is equivalent to $I(E(\ln Y_{iy}|\boldsymbol{x_{it}} = \boldsymbol{0}, j=1) \leq \cdots \leq E(\ln Y_{iy}|\boldsymbol{x_{it}} = \boldsymbol{0}, j=J))$.

Table 1. Summary Statistics

Variables	Mean	SD	Minimum	Maximum
$Y =$ Rice output (tons)	6.47	5.08	0.09	31.10
$X_1 =$ Area	2.12	1.45	0.20	7.00
$X_2 =$ Labor	107.20	76.65	8.00	436.00
$X_3 =$ Fertilizer	187.05	168.59	3.40	1,030.90

The current section is organized as follows: the application-specific prior parameter values are detailed, followed by the estimation results and the economic interpretations. Some common economic constraints incorporated in our models are also discussed. We then proceed to test the fit and robustness of our results.

Prior Parameter Values

We begin by selecting a value of $\mathbf{1}_J$ hyperparameter δ, which is a J-dimensional vector of ones, to reflect the least-possible prior knowledge about the probability occurrence of each state of nature. We select prior values for the slopes and intercepts, $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\kappa}$, so that the intercept terms of $\boldsymbol{\xi}$ comply with our labeling restriction and the slopes of the main variables have a relatively small value with a large variability, and thus are not significant a priori. We assign no prior significance and large variability to the coefficients corresponding to the interaction terms. We select precision-related hyperparameters, v, g, and m, based on the range of the errors resulting from a frequentist regression using a translog functional form and considering a linear time trend. We use the result for the number of states in O'Donnell and Griffiths, 3, as the prior value for λ, the mean of the distribution describing the number of states. Finally, we consider a prior median technical efficiency (TE) τ^* of 0.875 with a prior lower bound of 0.7 (see corresponding subsections for the sensitivity analysis performed on λ, τ^* and the prior lower bound of TE). Table 2 summarizes the prior values for the hyperparameters.

Table 2. Summary of Prior Values for Model Hyperparameters

Hyperparameter	Prior Value	Justification
Intercepts on $\boldsymbol{\zeta}, \boldsymbol{\kappa}$	$(2j\text{-}1)/2J^{\text{th}}$ percentile of output vector, where j is the state number, 2.25	Comply with labeling restriction. Assign large prior variance.
Time trend on $\boldsymbol{\zeta}, \boldsymbol{\kappa}$ (for linear time trend model)	0.02, 0.15	Annual output growth rate from database. Assign large prior variance.
Time dummies on $\boldsymbol{\zeta}, \boldsymbol{\kappa}$ (for dummy trend model)	0.02*(year-1990), 0.15	Annual output growth rate from database. Assign large prior variance.
Slopes on $\boldsymbol{\zeta}, \boldsymbol{\kappa}$	0.5, 6.5	Small value, not significant due to its large variance.
Interaction terms on $\boldsymbol{\zeta}, \boldsymbol{\kappa}$	0, 26	Consider no impact of input interactions a priori. Also, assign large prior variance.
δ	$\mathbf{1}_J$	Consider all states to be equiprobable a priori; assign smallest possible weight to prior distribution.
v g m	2 0.2 $100 * g/(v * (\max_{it} e_{it} - \min_{it} e_{it})^2)$	Same as Richardson and Green (1997) to ensure data-dependent vague precision prior; $\{e_{it}\}$ are the errors of a least squares regression with the translog function.
λ	3	Best estimation obtained by O'Donnell and Griffiths.
Max. of number of allowed states	100	Needed for BDMCMC; choose very large value so as not to affect estimation.
τ^*	0.875	Assumed prior median technical efficiency.
ς	0.7	Assumed prior lower bound for technical efficiency.

Posterior Parameter Estimates and Interpretation

We run both the linear and dummy time trend models, as well as their monotonicity-constrained versions. Here, we discuss the results of the monotonicity-constrained linear and dummy time trend models and compare them against those of the monotonicity-constrained fully state-contingent model in O'Donnell and Griffiths. We run 5,500 iterations for all our models, discarding the first 500 as burn-in. From the observed mixing on the trace plots for the number of states, precisions, state probabilities, and frontier coefficients (available in the online supplementary appendix for both the monotonicity-constrained dummy time trend model and the monotonicity-constrained linear time trend model), this period appears to be sufficient. After estimating the frontier, we find that the posterior distribution of the number of states of nature, J, is unimodal with a mode 3 for both monotonicity-constrained models. Moreover, the highest probability region of the posterior distribution of J allows us to create a 90% credible set including the values 2, 3, or 4 (figure 2). This distribution, which provides evidence of at least two major groups of farms based on differences in environmental conditions, supports our state-contingent hypothesis. Moreover, if hyperparameter Θ is held fixed, our linear time trend model is equivalent to the fully state-contingent model in O'Donnell and Griffiths, but with Θ held fixed, our estimation procedure results in a posterior distribution for the number of states with a mode of 2 versus the 3 states estimated in O'Donnell and Griffiths. This latter finding suggests that O'Donnell and Griffiths overestimate the number of states by relying on BIC as the model-selection criterion. Table 3 shows the results for our models for $J = 3$ and for the fully state-contingent models in O'Donnell and Griffiths.

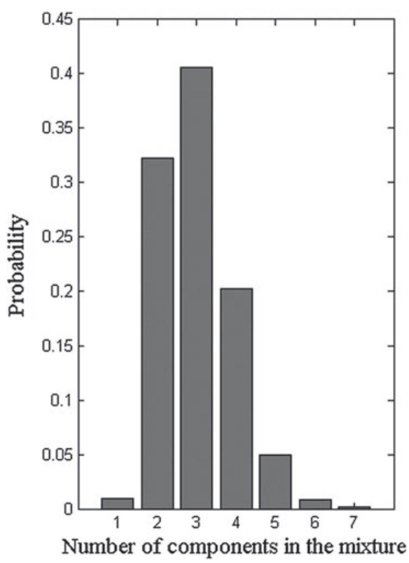

Figure 2. Posterior distribution for the number of components in the mixture for dummy time trend model

The expected efficient log-output for the first state (the intercept term) in our two models is not as low as the estimates in O'Donnell and Griffiths. Hence, we infer that factors other than land area, labor, and fertilizer have fewer detrimental effects relative to the results of O'Donnell and Griffiths. For instance, if we exponentiate the intercepts to obtain expected efficient yields for the first year of the timespan, 1990, we find that for state 2 of our dummy time trend model this figure is only 6% higher than that of the most unfavorable state (state 1), whereas this same comparison results in a 200% difference for O'Donnell and Griffiths.

We define four confidence levels (99%—highly; 95%—moderately; 90%—mildly; and 85%—barely significant)[3] to compare our two models with O'Donnell and Griffiths. For our two models and O'Donnell and Griffiths, the area, labor, and the state-contingent intercepts are significant at approximately the same levels for all states. The primary difference is in the estimated effect of fertilizer, which is highly significant for all states in our dummy model, moderately significant for two states in our linear model, and highly significant only for state 1 in O'Donnell and Griffiths; recall that state 1 has unfavorable (low expected yield) weather conditions. Unlike O'Donnell and Griffiths, however, our two models do not estimate negative fertilizer elasticities for state 1. O'Donnell and Griffiths posit excessive fertilizer application as detrimental to rice yield in bad weather conditions. We suggest that the disagreement

[3] For the sake of computational simplicity, we use classical hypothesis testing because we only want to evaluate the intercepts, time trends, and elasticities, which have a posterior normal distribution. Thus, the normality of classical hypothesis testing holds.

Table 3. Estimated Production Frontier Coefficient Means for BDMCMC Linear, Dummy Time Trend, and O'Donnell and Griffiths Models at the Posterior Mode $J = 3$

		Means					
		Linear Trend		Dummy Trend		O'Donnell and Griffiths SC-all	
Coefficient	State	Free	Mono	Free	Mono	Free	Mono
Intercept	1	1.886****	1.886****	1.917****	1.918****	1.118****	1.112****
	2	2.002****	2.002****	1.983****	1.976****	1.814****	1.803****
	3	2.098****	2.101****	2.061****	2.049****	2.082****	2.079****
Time Trend	1	0.024*	0.023*	Multiple	Multiple	0.028**	0.029**
	2	0.011	0.012	Multiple	Multiple	−0.014	−0.013
	3	0.013	0.014	Multiple	Multiple	0.009	0.010
ln(Area)	1	0.708****	0.696****	0.735****	0.623****	0.615***	0.434***
	2	0.645****	0.637****	0.66****	0.584****	0.133	0.143**
	3	0.556****	0.546****	0.586****	0.577****	0.561*	0.369***
ln(Labor)	1	0.016	0.036	−0.033	0.127	−0.333	0.107
	2	0.071	0.079	0.044	0.138	0.024	0.119**
	3	0.231**	0.241**	0.201	0.204**	−0.106	0.184**
ln(Fertilizer)	1	0.197	0.193	0.212	0.187***	−0.199	−0.368***
	2	0.151**	0.151**	0.183	0.18***	0.112	0.049
	3	0.15**	0.152**	0.159	0.169***	0.313**	0.231
ln(Area)^2/2	1	−0.259	−0.297	−0.256***	−0.297	0.011	0.006
	2	−0.558	−0.545	−0.40****	−0.398	−0.408	−0.395
	3	−0.637	−0.619	−0.59****	−0.567	−0.690	−0.158
ln(Area)* ln(Labor)	1	0.420	0.481	0.413	0.454	0.176	0.213
	2	0.611	0.592	0.494	0.504	0.336	0.391
	3	0.740	0.731	0.732	0.700	0.700	0.379
ln(Area)* ln(Fert.)	1	0.157	0.141	0.165	0.093	0.171	0.048
	2	0.159	0.169	0.156	0.106	−0.005	−0.048
	3	0.039	0.033	0.047	0.047	−0.529	−0.492
ln(Labor)^2/2	1	−0.441	−0.537	−0.442**	−0.399	−0.580	−0.366
	2	−0.493	−0.481	−0.496**	−0.501	−0.145	−0.066
	3	−0.772	−0.746	−0.827***	−0.779	−0.221	−0.408
ln(Labor)* ln(Fert.)	1	−0.307	−0.278	−0.287	−0.262	−0.288**	−0.223
	2	−0.349	−0.346	−0.269	−0.228	−0.181	−0.204
	3	−0.145	−0.153	−0.106	−0.111	0.555	0.637**
ln(Fert.)^2/2	1	0.151	0.133	0.119	0.136	−0.172	−0.228*
	2	0.129	0.117	0.085	0.091	0.152**	0.145**
	3	0.035	0.046	0.005	0.016	−0.395*	−0.385*
Precision	1	11.540	11.433	14.616	14.342	5.81	5.648
	2	12.809	12.985	14.752	14.701	8.513	8.528
	3	15.138	15.248	16.807	16.582	8.303	8.346
State Prob.	1	0.296	0.293	0.347	0.330	0.312	0.306
	2	0.345	0.344	0.325	0.335	0.363	0.364
	3	0.358	0.361	0.327	0.333	0.325	0.330

Note: Asterisks *, **, ***, and **** denote significance at the 15%, 10%, 5%, and 1% levels, respectively.

Table 3. (Continued). Estimated Production Frontier Coefficient Standard Deviations for BDMCMC Linear, Dummy Time Trend, and O'Donnell and Griffiths Models at the Posterior Mode $J = 3$

		Standard Deviations					
		Linear Trend		Dummy Trend		O'Donnell and Griffiths SC-all	
Coefficient	State	Free	Mono	Free	Mono	Free	Mono
Intercept	1	0.102	0.098	0.075	0.071	0.200	0.196
	2	0.079	0.077	0.070	0.066	0.087	0.087
	3	0.085	0.085	0.080	0.074	0.108	0.103
Time Trend	1	0.020	0.020	Multiple	Multiple	0.018	0.018
	2	0.015	0.015	Multiple	Multiple	0.016	0.016
	3	0.014	0.014	Multiple	Multiple	0.014	0.014
ln(Area)	1	0.270	0.268	0.060	0.158	0.296	0.239
	2	0.209	0.206	0.204	0.157	0.195	0.103
	3	0.192	0.192	0.230	0.161	0.471	0.241
ln(Labor)	1	0.268	0.265	0.215	0.109	0.242	0.094
	2	0.206	0.203	0.199	0.115	0.184	0.090
	3	0.173	0.171	0.188	0.131	0.347	0.144
ln(Fertilizer)	1	0.174	0.177	0.218	0.105	0.239	0.227
	2	0.122	0.119	0.197	0.096	0.124	0.110
	3	0.102	0.108	0.170	0.090	0.201	0.190
ln(Area)ˆ2/2	1	0.742	0.744	0.128	0.641	0.339	0.344
	2	0.738	0.688	0.113	0.674	0.933	0.927
	3	0.921	0.867	0.100	0.837	1.795	1.416
ln(Area)* ln(Labor)	1	0.659	0.647	0.602	0.567	0.341	0.334
	2	0.656	0.636	0.709	0.607	0.647	0.641
	3	0.780	0.764	0.905	0.770	1.237	1.154
ln(Area)* ln(Fert)	1	0.414	0.407	0.517	0.330	0.241	0.229
	2	0.380	0.360	0.633	0.341	0.456	0.415
	3	0.430	0.407	0.815	0.401	0.581	0.578
ln(Labor)ˆ2/2	1	0.882	0.896	0.327	0.786	0.510	0.489
	2	0.861	0.864	0.361	0.811	0.733	0.729
	3	0.993	0.981	0.417	0.975	1.332	1.296
ln(Labor)* ln(Fertilizer)	1	0.374	0.399	0.735	0.301	0.223	0.218
	2	0.365	0.354	0.837	0.327	0.459	0.437
	3	0.404	0.391	0.993	0.372	0.495	0.484
ln(Fertilizer)ˆ2/2	1	0.281	0.288	0.297	0.205	0.189	0.186
	2	0.241	0.232	0.351	0.235	0.109	0.106
	3	0.297	0.280	0.390	0.267	0.358	0.357
Precision	1	3.075	3.014	3.555	3.870	1.082	1.084
	2	3.221	3.254	3.866	4.415	1.298	1.286
	3	3.639	3.706	4.494	0.094	1.466	1.513
State Prob.	1	0.091	0.091	0.087	0.097	0.060	0.063
	2	0.107	0.109	0.092	0.087	0.050	0.051
	3	0.093	0.091	0.081	0.060	0.058	0.062

Note: Asterisks *, **, ***, and **** denote significance at the 15%, 10%, 5%, and 1% levels, respectively.

in our parameter estimates and O'Donnell and Griffiths' are likely driven by the differences between our worst state's expected yield results. O'Donnell and Griffiths' worst state is associated with a significantly lower yield compared to ours (see table 3). O'Donnell and Griffiths' significantly lower estimates of the yield in state 1 are likely to have effects on the parameter estimates of the production function associated with that state. Our positive elasticity results for fertilizer are supported by SriRamaratnam et al. (1987). These authors show that for yield per-unit area levels similar to the ones obtained in our state 1, fertilizer ranging from very low to very high concentrations per unit of area has a positive effect on field yield. Griffin et al. (1985) show similar results in the context of ratoon crops. Thus, we conclude that for all states estimated in our application, fertilizer has a non-decreasing relationship to yield. Furthermore, this result does not contradict the notion of fertilizer being a risky input in a profit-maximization context, as the increase in yield may not outweigh its acquisition and application costs. Moreover, O'Donnell and Griffiths show moderately or barely significant elasticities for the squared fertilizer term for all states. Combining these with their fertilizer elasticities, O'Donnell and Griffiths' results and ours are roughly consistent for states 2 and 3, suggesting a moderately to barely significant positive elasticity of fertilizer on rice yields.

The planted area elasticities of our two models seem to be slightly higher for states 1 and 3 and significantly higher for state 2, suggesting a higher per hectare yield relative to O'Donnell and Griffiths' results. For labor, our estimated elasticities and O'Donnell and Griffiths are similar. Finally, and unlike O'Donnell and Griffiths, our two models show no significant second-order or interaction terms, and first-order input coefficients have a sum close to unity. From these results, we conclude that approximately constant returns-to-scale characterize the Tarlac region's rice production.

The time trend effects are only significant for state 1 in O'Donnell and Griffiths' results, indicating either technological progress regarding the methods to handle bad weather conditions, or increasingly more benign "bad" weather conditions throughout the time span. Our state-contingent linear trend model, which shows that the time trend coefficients are barely significant for state 1, partially supports this insight. Our time trend coefficient for state 3, which is the largest in magnitude, aligns with O'Donnell and Griffiths. Furthermore, our dummy time trend model provides additional insight into changes in the production environment of the Tarlac region over time. Thus far, the only indicator of the overall change in production conditions over time is O'Donnell and Griffiths' model with J=1 (equivalent to the RE model in the O'Donnell and Griffiths paper), which indicates a significant linear trend coefficient of 0.014. Figure 3 shows a more complicated pattern of output fluctuation over time, as the dummy time trend model's coefficient values relative to the base year are at least barely significant for every year except for 1991 and 1994, and they show that both "good" (1997) and "bad" (1996) years occur.

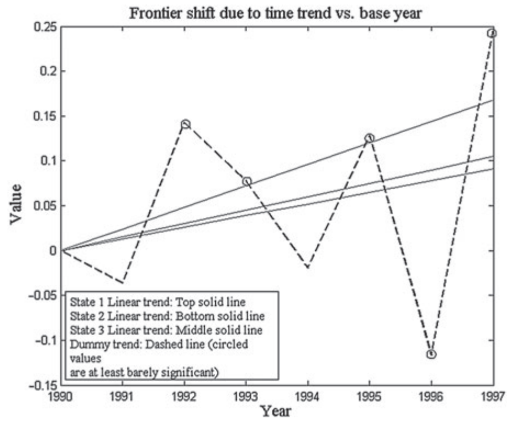

Figure 3. Dummy time trend and state-contingent linear time trend comparison

The state-contingent intercepts for the dummy time trend model are slightly more similar to one another than for the linear time trend model, while the estimated precisions are larger for the dummy trend model. Both of these results indicate that the dummy trends capture a larger percentage of output variability. This finding translates into a smaller percentage of the expected output being assigned to the unobserved variables modeling the different states (for details, see the label-switching subsection). In general, the precisions of the noise terms are larger in our two models than in O'Donnell and Griffiths, which is consistent with the smaller MSE obtained using our models (for details, see the goodness-of-fit subsection).

The technical efficiency estimates for our two models have a Spearman correlation coefficient of roughly 80% when compared to O'Donnell and Griffiths, meaning that the efficiency rankings do not change radically (see figure 5). For the 44 rice farms as a group, the mean technical efficiency estimated using either the dummy or the linear time trend models is lower than O'Donnell and Griffiths. However, if we set a lower bound on technical efficiency to 0.8, our result is less than 1% different from that of O'Donnell and Griffiths. In any case, O'Donnell and Griffiths' TE *distribution* differs from the results of our two models, based on a series of Kolmogorov-Smirnov tests with p-values lower than 0.1% for the null hypothesis of distributional equality. One exception occurs when we compare O'Donnell and Griffiths' TE distribution with our linear model with a lower bound on technical efficiency set to 0.8, where the p-value for the null hypothesis is equal to 0.105. Hence, the hypothesis is not rejected at the usual significance levels, indicating statistically equivalent distributions.

When comparing the random effects model against a fully state-contingent model, O'Donnell and Griffiths find a 7% higher mean TE when using the state-contingent model. While we can neither validate nor refute O'Donnell and Griffiths' result, since our mean TE depends on the prior lower bound, we can assess the degree to which mean TE increases when using a state-contingent frontier by estimating a random effects model with the same value for the lower bound on TE. For the dummy time trend model, a less than 1% increase in the TE is observed when considering a state-contingent frontier instead of a random effects frontier, suggesting that state-contingency does not account for as much variability in output as O'Donnell and Griffiths' results indicate.

Incorporating Monotonicity and Convexity Constraints into the Frontier Model

To impose monotonicity of inputs, we use an accept-reject method to sample from the truncation region for the multivariate normal draws of $\boldsymbol{\beta}$ (area, labor, fertilizer). We define the truncated region by the monotonicity and labeling restrictions. Our BDMCMC algorithm produces satisfactory monotonicity-constrained draws for both the linear time trend and the dummy time trend models. Table 3 shows that there is not a large difference between the restricted and unrestricted models, implying that the unrestricted models nearly satisfy monotonicity. Regarding the imposition of a convexity constraint, we form the Hessian matrix to test quasi-convexity in the inputs used as suggested by O'Donnell and Coelli (2005). Given the selected dataset, production frontier model, and the distributional assumptions on the parameters, some observations do not comply with a convexity assumption. Hence, we conclude that by using a simple accept-reject method, it is not possible to estimate a convexity-constrained version of our two models using our BDMCMC algorithm for this data set.

Goodness-of-fit

To assess the goodness-of-fit of both our linear time trend and dummy time trend models, we compare the MSEs and the quantiles of the posterior distribution of the observed residuals with those of a Gaussian mixture distribution. First, we generate predicted output values from our state-contingent models by simulating from the posterior distribution of the number of states and then simulating from the posterior distribution of π, given the simulated value of J. Drawing the values for these parameters tells us both the state to which an observation is assigned and which production function will predict the output level. Second, we compare these predicted output levels to the observed output level to obtain a prediction error and to calculate MSE. Third, we compare the prediction error quantiles against those of a Gaussian mixture error with a mean of zero and the corresponding state-specific precisions to validate the distribution of the observed prediction errors. Fourth, since both the observed and theoretical standard errors depend on the generated values of the uniform random numbers we use to draw from the aforementioned posterior distributions, we run the error-generating algorithm until both the observed and theoretical standard error vectors reach stationarity.

We consider two simulation scenarios: the full posterior scenario, which weights predictions from all values of J using the full distribution for the number of states described previously, and the posterior mode scenario, which gives full weighting to the mode of the posterior number of states. Table 4 lists the MSE values obtained for the monotonicity constrained linear and dummy time trend models, and shows that the RMSE percentage is practically identical for all of our models and scenarios, and is below 5%, indicating a good in-sample performance of our estimated state-contingent production frontier model. Figure 4 and the online supplementary appendix show the QQ plots for the stationary vectors for the dummy time trend model and the linear time trend model, respectively. The full posterior scenario exhibits a slightly better fit, and is consistent with our conclusions using MSE. Moreover, the two models predict

Table 4. MSE, RMSE, %RMSE Comparison Between Different Models

MSE	Linear	Dummy	ODG	RMSE	Linear	Dummy	ODG	%RMSE	Linear	Dummy	ODG
$J=3$	0.075	0.074	0.290	$J=3$	0.27363	0.27161	0.53833	$J=3$	4.69%	4.65%	9.21%
Full	0.075	0.072	–	Full	0.2733	0.26925	–	Full	4.68%	4.61%	–

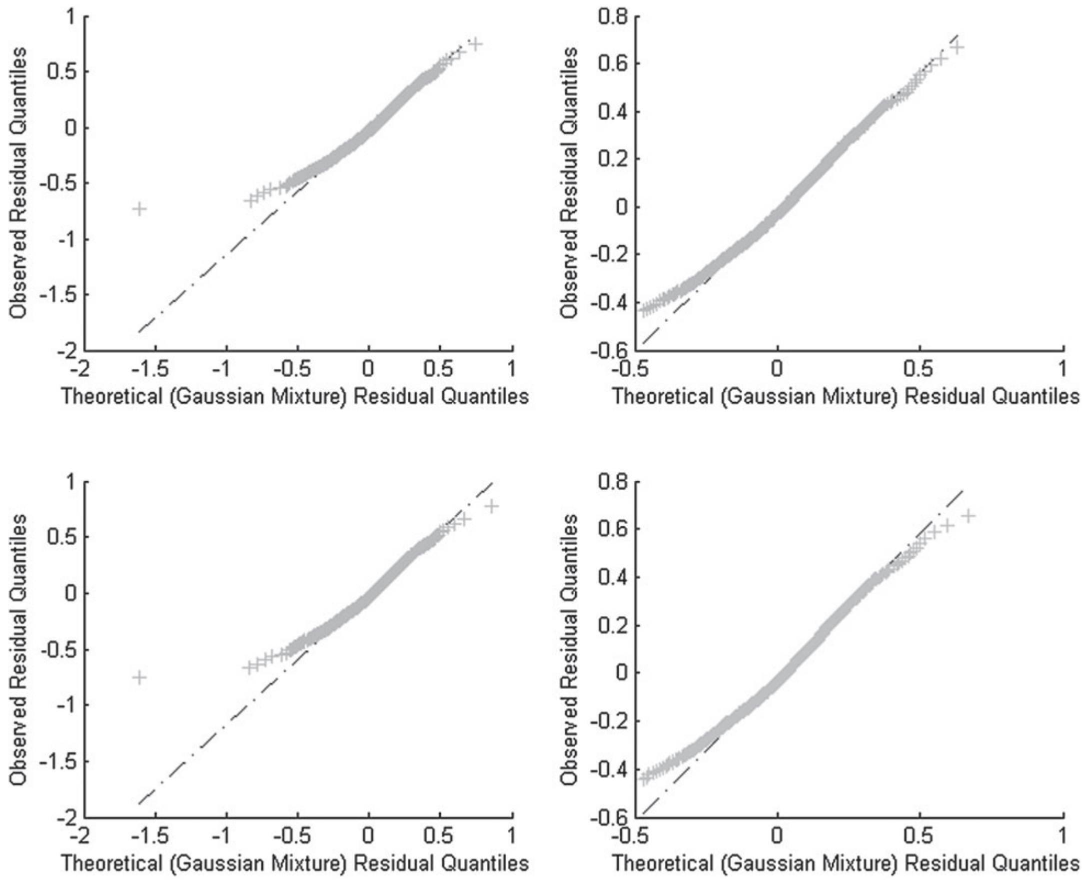

Figure 4. Goodness-of-fit results for dummy model

Note: Results are for the full posterior scenario (top-left panel); 4% outliers removed for full posterior scenario (top-right panel); results for mode scenario (bottom-left panel); 4% outliers removed for mode scenario (bottom-right panel).

different sets of outliers, each containing 15 observations, or roughly 4% of the total sample (the right column of figure 4 shows the outlier-free fit results). Based on the minimal difference in MSE obtained when using the mode for J instead of its full distribution, we conclude: *a*) for our application, using the mode of the number of states to predict efficiency provides a reasonable fit; and *b*) re-running the model assuming the mode of the posterior distribution of J is the correct number of states gives a parsimonious and nearly as well-fitting model (bottom panels of figure 4). Our conclusions suggest that, in this particular application, using the mode of the posterior distribution for the number of states is a reasonable choice due to the small MSE difference versus using the full posterior distribution, although this may not be the case in a general setting. Comparing both scenarios' goodness-of-fit shows that using a model based on a point estimate of the number of states, such as using the mode of the posterior distribution or the BIC-minimizing number of states, is in fact a special case of the full posterior model we develop. Table 4 also compares the MSE figures from our models and O'Donnell and Griffiths. The

Table 5. Sensitivity Analysis on Hyperparameter λ, Prior Mean on the Number of States for the Dummy Time Trend Model

Prior lambda	Posterior mode for J	Posterior mean for J	$P(J=2\|X)$	$P(J=3\|X)$	90% HPD[a]
1	3	3.02	0.3162	0.4076	{2,3,4}
2	3	3.07	0.2938	0.413	{2,3,4}
3	3	3.07	0.306	0.4018	{2,3,4}
4	3	3.1	0.285	0.395	{2,3,4}
5	3	3.1	0.3056	0.386	{2,3,4}

Note: Superscript[a] indicates the Highest Posterior Density set.

MSE obtained from our models is roughly $\frac{1}{4}$ the magnitude of the MSE obtained by using O'Donnell and Griffiths' estimated frontier and could be a result of the additional smoothing provided in the estimation of the precision parameters, meaning that our models provide a better explanation of the variability in output.

Sensitivity Analysis

Implementing the BDMCMC algorithm allows us to analyze the estimated posterior probability distribution for the number of states of nature. To determine the model's robustness, we investigate whether our results depend on the mean value of the prior distribution on the number of states of nature, λ. We assign integer values ranging from 1 to 5 for lambda. In all instances, the posterior shows evidence of at least 2 states.

For both the linear and dummy-trend models, the posterior mean and mode for J indicate that $J=3$ in most cases. The 90% Highest Posterior Density (HPD) set for the value of J seems to be invariant over the distinct choices for λ. Finally, the maximum number of possible states $J=100$ is far from being reached during the estimation process, since 8 is the highest number of components at any point of time for all choices of λ. Therefore, we conclude that this parameter is immaterial to the estimation of the model. Table 5 summarizes the findings for the dummy time trend model.

To conduct sensitivity analysis on our assumptions about the inefficiency levels, we vary the prior lower bound on TE from its base value of 0.7 on the range [0.5, 0.8] in 0.1 increments. Figure 6 shows that the mean of the posterior distribution for TE shifts upward relative to the prior lower bound, and that the distribution changes to accommodate the smaller range of TE. Figure 5, a

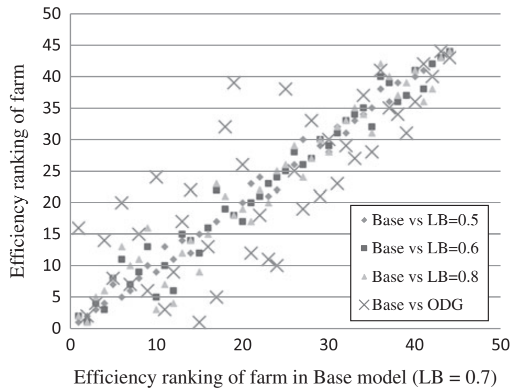

Figure 5. Rankings of the three alternative prior TE lower bound assumptions against the base value of 0.7 for the dummy time trend model

scatterplot of rankings considering all sensitivity scenarios against our base assumption, shows that there are no significant changes in the inefficiency rankings of the firms. Computing Spearman's correlation coefficients for the base scenario rankings against those of the three alternative lower bound values shows that the lowest coefficient is 96.4% for the dummy time trend model, which still indicates a strong relationship between the rankings. Thus, we conclude that while the relative efficiency rankings do not change depending on the value of the lower bound on TE, our inefficiency distribution is sensitive to this assumption.

Finally, we also perform sensitivity analysis on the prior median level of TE. We consider three different values for τ*: a low value of 0.825, O'Donnell and Griffiths base value of 0.875, and a high value of 0.925. Figure 6 shows that the inefficiency distribution is approximately independent on the choice of τ* for our dummy time trend model. The efficiency rankings are approximately maintained for the three scenarios, with the lowest

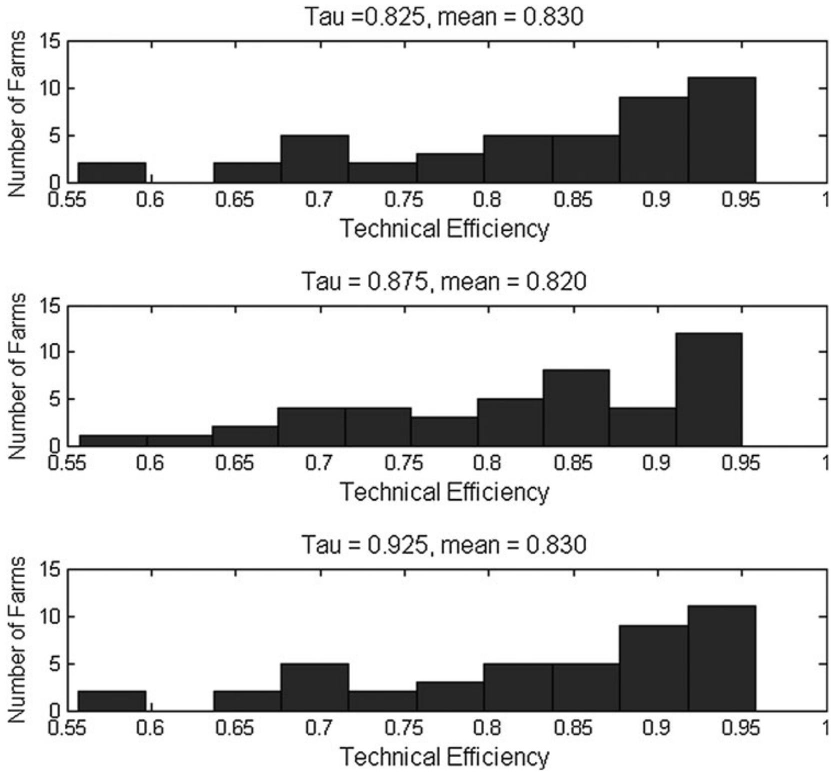

Figure 6. Technical efficiency distribution for low (top panel), base (middle panel), and high (bottom panel) values of for the dummy trend model

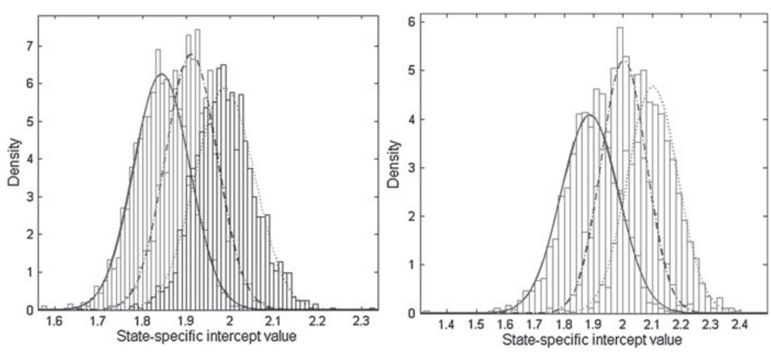

Figure 7. Distribution of intercept (expected output at time 0) for three states, the mode of J. Fitted curves for dummy time trend model (left panel), linear time trend model results (right panel). Solid line denotes distribution for state 1 intercept, dashed line denotes distribution for state 2 intercept, and dotted line denotes distribution for state 3 intercept.

Spearman's correlation coefficient being greater than 99% overall compared to the base scenario. The results for our linear time trend model for all the sensitivity analyses performed are similar and their details can be found on the online supplementary appendix.

Label-switching

Our labeling restriction alternative addresses the mixture component label-switching problem by setting a labeling restriction to ensure that the drawn intercept value from a lower-indexed state has a lower value than that of a higher-indexed state on each iteration of the BDMCMC algorithm. Since the drawn values differ for each iteration, varying degrees of overlap between the intercept distributions can exist. Figure 7 (left panel) shows that the state-contingent linear time trend model gives a large degree of overlap between these distributions, that is, the distribution of the first state's intercept has approximately a 50% overlap with the second state's intercept, and the second state's intercept distribution has almost a 55% overlap with the third state's intercept distribution. Figure 7 (right panel) shows that the overlap is slightly more significant for the dummy trend model, that is, if we plot all of the intercept distributions together, the compound histogram is slightly closer to unimodality, making the state-contingency of the intercepts less clear. In other words, the level of state-contingency of mean log-output diminishes only slightly when using a more flexible model to capture the time shifting effect, and continues to support the state-contingent hypothesis about the mean yield per state. Moreover, some coefficients such as the labor-related components are noticeably state-contingent when comparing states 1 and 3 (see the dummy time trend model results in table 3).

Conclusions

This article describes a new BDMCMC algorithm that efficiently estimates the number of states in a state-contingent model. This approach improves upon the Bayesian estimation of state-contingent models because the posterior distribution of this parameter can be visualized, thus providing more insights into the nature of the unobserved variables generating the states. After only one run of the model there is enough information to weight the outputs obtained for models with different numbers of states in a straightforward manner. Computing a goodness-of-fit analysis for a state-contingent model of production allows us to determine, by assessing the difference in MSE of both scenarios, whether the full posterior distribution of the number of states enhances the model significantly versus using only its mode. The experimental results derived from a case study of 44 rice farms in the Tarlac region of the Philippines shows an insignificant difference between using the mode of the posterior distribution versus using the complete distribution. Utilizing a state-independent dummy time trend, we estimate the differences in mean output levels across states to be slightly smaller than O'Donnell and Griffiths' estimates.

Our finding that a state-contingent linear time trend could not explain the complex time-shifting effect of the frontier suggests that changing weather patterns from year to year have a non-linear effect on output. The similarity in the mean output levels of states in our dummy time trend model suggests that bad/good years affect the rice-producing region more uniformly than indicated by O'Donnell and Griffiths. Nevertheless, evidence of state-contingency can still be argued by the differences in labor elasticities we find for states 1 and 3, as well as the slight differences in the state-contingent mean output. The unimodal posterior distribution on the number of states indicates that the interactions between unobserved variables are complex and probably interdependent. We suggest that the inability to impose a convexity constraint could be due to the limited flexibility of the parametric framework, distributional assumptions, or the presence of outliers in the dataset. A complementary or alternative explanation for the inability to impose convexity could also relate to the significant yearly shifts in the observed log-output. For example, in 1996 there is a distinct possibility that the group of farms as a whole was on a non-convex portion of a classical s-shaped yield curve since the mean log-output in 1996 was well below the average of the full time span. Comparing the results of our two models shows that our efficiency rankings are roughly consistent with O'Donnell and Griffiths, and that the main benefit of the O'Donnell and Griffiths model, the ability to avoid misinterpreting state-contingency as inefficiency, is maintained. In terms of input elasticities, our area and labor elasticities are similar to O'Donnell and Griffiths, although we estimate a positive fertilizer elasticity for our lowest output state, whereas O'Donnell and Griffiths find a negative effect of fertilizer. The difference is driven by the higher average yield estimates in state 1 that we obtain compared to O'Donnell and Griffiths. Our positive fertilizer elasticities, given our estimated yield per unit area in that state, are consistent with the previous rice crop literature. Finally, our models provide slightly larger yield estimates and exhibit a significantly lower MSE.

Further work remains to be done regarding the estimation of state-contingent frontiers. A key limitation is the challenge of imposing a convexity constraint. This is likely an issue related to this data set. However, convexity could be imposed more easily if we used a non-parametric regression method such as MBCR (Hannah and Dunson 2011). Also, our current model needs a pre-specified lower bound on technical efficiency to be estimated. While we have performed a sensitivity analysis on the related parameter, the approach in Mahendran et al. (2012) could be used to select this parameter optimally in terms of a specific criterion such as the MSE.

Supplementary Material

Supplementary online appendix is available at http://oxfordjournals.org/our_journals/ajae/online.

References

Baltagi, B.H., and J.M. Griffin. 1988. A General Index of Technical Change. *The Journal of Political Economy* 96: 20–41.

Banfield, J.D., and A.E. Raftery. 1993. Model-based Gaussian and Non-Gaussian Clustering. *Biometrics* 49 (3): 803–21.

Biernacki C., G. Celeux, and G. Govaert. 1998. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. Technical Report No. 3521, Institut National de Recherche en Informatique et en Automatique, Rhône-Alpes.

———. 2000. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (7): 719–25.

Biernacki, C., and G. Govaert. 1997. Using the Classification Likelihood to Choose the Number of Clusters. *Computing Science and Statistics* 29 (2): 451–57.

Biller, C. 2000. Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models. *Journal of Computational and Graphical Statistics* 9 (1): 122–40.

Chavas, J.P. 2008. A Cost Approach to Economic Analysis under State-contingent Production Uncertainty. *American Journal of Agricultural Economics* 90: 435–46.

Chopin, N., and F. Pelgrin. 2004. Bayesian Inference and State Number Determination for Hidden Markov Models: An Application to the Information Content of the Yield Curve About Inflation. *Journal of Econometrics* 123 (2): 327–44.

Denison, D.G.T., Mallick, B.K., and A.F.M. Smith. 1998. Automatic Bayesian Curve Fitting. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 60 (2): 333–50.

Gelfand, A.E., and A. Kottas. 2001. Bayesian Semiparametric Median Regression Modeling. *Journal American Statistical Association* 96 (4): 1458–68.

Geweke, J. 1991. Efficient Simulation from the Multivariate Normal and Student t-Distributions Subject to Linear Constraints. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E.M. Keramidas and S.M. Kaufman, 571–78. Seattle: Interface Foundation of America.

Green, P.J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* 82 (4): 711–32.

Griffin, R.C., M.E. Rister, J.M. Montgomery, and F.T. Turner. 1985. Scheduling Inputs With Production Functions: Optimal Nitrogen Programs for Rice. *Southern Journal of Agricultural Economics* 17 (1): 159–68.

Hannah, L., and D. Dunson. 2011. Bayesian Nonparametric Multivariate Convex Regression. Technical Report arXiv: 1109.0322v1. Department of Statistical Science, Duke University, Durham, North Carolina.

Hurn, M., A. Justel, and C.P. Robert. 2003. Estimating Mixtures of Regressions. *Journal of Computational and Graphical Statistics* 12 (1): 1–25.

Mahendran, N., Z. Wang, F. Hamze, and N. de Freitas. 2012. Adaptive MCMC with Bayesian Optimization. *Journal of Machine Learning Research – Proceedings Track* 22: 751–60.

Nauges, C., C.J. O'Donnell, and J. Quiggin. 2011. Uncertainty and Technical Efficiency in Finnish Agriculture: a State-contingent Approach. *European Review of Agricultural Economics* 38 (4): 449–67.

O'Donnell, C.J., and T.J. Coelli. 2005. A Bayesian Approach to Imposing Curvature on Distance Functions. *Journal of Econometrics* 126 (2): 493–523.

O'Donnell, C.J., and W.E. Griffiths. 2006. Estimating State-contingent Production Frontiers. *American Journal of Agricultural Economics* 88 (1): 249–66.

O'Donnell, C.J., and S. Shankar. 2010. Regulatory Reform and Economic Performance in U.S. Electricity Generation Estimating State-allocable Production Technologies When there are Two States of Nature and State Allocations of Inputs are Unobserved. CEPA Working Papers Series WP042010, School of Economics, University of Queensland, Australia.

Orea, L., and S.C. Kumbhakar. 2004. Efficiency Measurement Using a Latent Class Stochastic Frontier Model. *Empirical Economics* 29 (1): 169–83.

Quiggin, J.C., and R.G. Chambers. 2006. The State-contingent Approach to Production under Uncertainty. *Australian Journal of Agricultural and Resource Economics* 50 (2): 153–69.

Richardson, S., and P.J. Green. 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 (4): 731–92.

Robert, C.P. 1995. Simulation of Truncated Normal Variables. *Statistics and Computing* 5 (2): 121–25.

Serra, T., S. Stefanou, and A.O. Lansink. 2010. Dynamic State Contingent Production and Capital Accumulation in U.S. Agriculture. *European Review of Agricultural Economics* 37 (3): 293–312.

Sri Ramaratnam, S., D.A. Bessler, M.E. Rister, J. Novak, and J.E. Matocha. 1987. Fertilization under Uncertainty: An Analysis Based on Subjective Producer Yield Expectations. *American Journal of Agricultural Economics* 69 (2): 349–57.

Stephens, M. 2000. Bayesian Analysis of Mixtures with an Unknown Number of Components – an Alternative to Reversible Jump Methods. *Annals of Statistics* 28 (1): 40–74.

Villano, R.A., C.J. O'Donnell, and G.E. Battese. 2004. An Investigation of Production Risk, Risk Preferences and Technical Efficiency: Evidence from Rainfed Lowland Rice Farms in the Philippines. Article presented at AsiaPacific Productivity Conference, Brisbane, Australia.