TOPICS IN IDENTIFICATION, LIMITED DEPENDENT VARIABLES, PARTIAL OBSERVABILITY, EXPERIMENTATION, AND FLEXIBLE MODELING: PART B

ADVANCES IN ECONOMETRICS

Series Editors: Thomas B. Fomby,
R. Carter Hill, Ivan Jeliazkov,
Juan Carlos Escanciano,
Eric Hillebrand, Daniel L. Millimet,
Rodney Strachan,
David T. Jacho-Chávez and
Alicia Rambaldi

Previous Volumes:

Volume 29: Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill

Volume 30: 30th Anniversary Edition – Edited by Dek Terrell and Daniel Millmet.

Volume 31: Structural Econometric Models – Edited by Eugene Choo and Matthew Shum.

Volume 32: VAR Models in Macroeconomics — New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy

Volume 33: Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park

Volume 34: Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier.

Volume 35: Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman

Volume 36: Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee.

Volume 37: Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage, and R. Kelley Pace

Volume 38: Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano

Volume 39: The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi

Volume 40A: Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A – Edited by Ivan Jeliazkov and Justin L. Tobias

ADVANCES IN ECONOMETRICS VOLUME 40, PART B

TOPICS IN IDENTIFICATION, LIMITED DEPENDENT VARIABLES, PARTIAL OBSERVABILITY, EXPERIMENTATION, AND FLEXIBLE MODELING: PART B

EDITED BY

IVAN JELIAZKOV
University of California, USA

JUSTIN L. TOBIAS
Purdue University, USA

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83867-420-5 (Print)
ISBN: 978-1-83867-419-9 (Online)
ISBN: 978-1-83867-421-2 (Epub)

ISSN: 0731-9053 (Series)

ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

CONTENTS

FOREWORD TO PART B

Part B of *Advances in Econometrics, Volume 40*, contains 11 chapters on topics that are distinct from those considered in Part A. Part B examines innovations in stochastic frontier (SF) analysis, nonparametric and semiparametric modeling and estimation, A/B experiments, and quantile regression. Together, the two parts form one of the largest volumes to date in the *Advances in Econometrics* series.

Two chapters in the volume address issues in SF models. The chapter by *Gholamreza Hajargasht and William E. Griffiths* considers a semiparametric SF model with correlated effects, where covariates in the production frontier enter nonparametrically. Specifically, they tackle a model of the form:

$$y_{it} = f(x_{it}) - u_i + v_{it},$$

where u_i is a time-invariant one-sided inefficiency term and f is modeled via penalized splines. They consider cases where x is both univariate and multivariate. For the multivariate case, they take up partially linear, additive, and general nonparametric forms, treating the final case via multivariate spline bases. The methodology is illustrated in extensive Monte Carlo experiments. *Eri Nakamura, Takuya Urakami, and Kazuhiko Kakamu* consider a fully parametric, cross-sectional SF model with a greater focus on a specific application. They apply the SF model to examine how the division of labor (as measured by the number of firm sections) impacts total costs among a sample of 79 Japanese water suppliers in 2010. They leverage the population served by the ith supplier as a source of exogenous variation and describe inference via Markov chain Monte Carlo (MCMC). They find that ignoring the endogeneity of the number sections (or functional units) within a supplying firm provides a conservative underestimate of its effect on total cost, and that costs rise with the number of sections. Their work points toward a policy implication where the integration of internal divisions within Japanese water suppliers leads to more cost-effective production.

A set of three chapters examines topics in semiparametric and nonparametric modeling. A chapter by *Justin L. Tobias and Joshua C. C. Chan* presents an interesting modeling approach that aims to resolve an important difficulty in the modeling of unknown regression functions. Because such modeling typically nests the important linear case at the boundary of the parameter space, it can only approximate, but not reproduce exactly, the linear regression model. The chapter proposes a hierarchical setup in which a Bernoulli variable determines whether the model is exactly linear or nonlinear (for some predetermined thresh-

old level of nonlinearity), and hence in the course of sampling, the realizations of that variable can be used to infer the probability of each of the two cases. The methodology and its usefulness are demonstrated in generated data experiments and an application involving the impact of the body mass index on labor market earnings. *Jianghao Chu, Tae-Hwy Lee, and Aman Ullah* consider the analysis of semiparametric single index models in big data settings with many regressors, of which few may be relevant. For this reason, variable selection is a key concern in this setting. The authors propose the SIM-Rodeo algorithm, which handles variable selection as local bandwith selection, and shows its consistency in variable selection. A Monte Carlo study shows its superior performance relative to alternatives such as the SIM-Lasso. In an important chapter, *Edward George, Purushottam Laud, Brent Logan, Robert McCulloch, and Rodney Sparapani* pursue a fully nonparametric extension of the Bayesian Additive Regression Tree (BART) model. While BART achieves flexible modeling of the conditional mean function through ensambles of trees, the authors generalize its applicablity to the case of non-Gaussian errors by using Dirichlet process mixtures (DPM). An important motivation for the chapter is that estimation should be not only pursued without relying on strong assumptions but also in a fairly automatic way that does not require too much tuning in a wide variety of settings. The authors show that the resulting DPMBART model performs well in simulated and real data studies with normal and non-normal errors.

This volume also contains two chapters developing methodologies useful for making quick and reliable inference in so-called A/B experiments. These experiments often involve millions of observations and are conducted routinely by large online businesses. The goal of these experiments is to assess the effect of various treatments such as changes in webpage design on revenue; the difference between treated and untreated outcomes is commonly referred to as *lift*. *John Geweke* introduces BABI, or Bayesian analysis of A/B data. He accounts for the prevalence of zeros and heavy tails in such data by considering a zero-inflated log-normal model. A key advantage of BABI is its ability to allow for informed priors surrounding values of lift; Geweke specifically considers nine states related to lift *breadth* and *depth* corresponding to either no effect, the anticipated effect, or some other effect of treatment in each dimension. He applies this methodology using data from 21 different experiments conducted by a business with a large online presence, and finds that BABI yields tighter credible intervals than alternate approaches. *Hedibert Freitas Lopes, Matthew Taddy, and Matthew Gardner* consider the issue of inference of means of heavy-tailed distributions, a feature that often characterizes distributions of online expenditure data. Their approach is to specify the likelihood as multinomial with a finite (but large) number of support points for values less than some threshold and to allow for heavy tails by specifying the density above the threshold as a generalized Pareto distribution. They show how Bayesian inference can be conducted on the sampling weights and tail parameters using a novel independence Metropolis–Hastings algorithm. The results are applied to a large A/B experiment containing more than 10^7 observations in eBay.

Three chapters in the volume relate to quantile regression. The chapter by *Mohammad Arshad Rahman and Angela Vossmeyer* discusses Bayesian techniques for estimating quantile regressions in binary longitudinal data settings. In Bayesian approaches, quantile regression typically employs the asymmetric Laplace distribution, noting the connection between the posterior mode under such a likelihood and a flat, improper prior, and the traditional frequentist quantile regression estimator. A useful additive mixture representation of the asymmetric Laplace, first noted by Kozumi and Kobayashi (2011) and skillfully adapted by Rahman and Vossmeyer to a binary longitudinal specification, yields computationally convenient MCMC inference. Rahman and Vossmeyer apply their methodology in two interesting applications: one involving determinants of female labor force participation and a second involving home ownership. *Debajit Dutta, Subhra Sankar Dhar and Amit Mitra* also consider the issue of quantile estimation, but of a location parameter in a stochastic volatility model. They derive the asymptotic distribution of the quantile estimator without assuming the density of the error is positive near the population quantile. They also discuss a Bayesian estimator based on the asymmetric Laplace likelihood. The third chapter in the set – by *Mohammad Arshad Rahman and Shubham Karnawat* – introduces a framework for flexible Bayesian quantile regression for ordinal outcomes. To overcome the problem that the skewness of the asymmetric Laplace distribution is determined by the chosen quantile, the authors study the generalized asymmetric Laplace (GAL) distribution. Their derivation of the GAL cumulative distribution and moment generating functions allows them to construct a likelihood function and posterior density that is explored by MCMC methods. The authors further demonstrate the advantages of their approach in an extensive simulation study and an application to public opionion on home ownership in the United States after the Great Recession.

The volume closes with "A Reaction," a brief comment from Dale J. Poirier on the June 2018 Advances in Econometrics Conference. Readers may take away from this short note a belief that the market for his classic *Intermediate Statistics and Econometrics* text has been surprisingly limited, and consumers of *Men Are from Mars, Women Are from Venus* should find his work equally appealing. We have all heard stories of successful marriages between devout Democrats and staunch Republicans, cat lovers and dog enthusiasts, and Michigan fans and Ohio State devotees. In two former students and contributors to this volume, Dale notes a similar warm stasis between a Bayesian and a frequentist. Whether or not the balanced presentation of his text served as a model for this otherwise inexplicably successful union of Hatfields and McCoys, the editors cannot say. We are tempted to recommend that Dale stick with his day job rather than branch out into couples therapy, but we also understand that his days are now free and unclaimed by the academy. Regardless of its true scope of influence, we remain grateful for his book and its many lessons, this opportunity to pay tribute to Dale, and the very positive impacts he has had on our careers.

This page intentionally left blank

ACKNOWLEDGMENTS

We would, once again, like to thank all contributing authors, conference presenters, and referees. We would also like to thank Nick Wolterman and Charlotte Wilson at Emerald Publishing for their assistance and guidance. Finally, we are very grateful for financial support from Emerald Publishing, Eviews, Orion Construction, the International Association for Applied Econometrics, UC Irvine's Department of Economics, School of Social Sciences and Donald Bren School of Information and Computer Science, and the Purdue University Research Center in Economics.

Ivan Jeliazkov
Justin L. Tobias

This page intentionally left blank

A SEMIPARAMETRIC STOCHASTIC FRONTIER MODEL WITH CORRELATED EFFECTS

Gholamreza Hajargasht[a] and William E. Griffiths[b]

[a]*Swinburne Business School, Australia*
[b]*University of Melbourne, Australia*

ABSTRACT

We consider a semiparametric panel stochastic frontier model where one-sided firm effects representing inefficiencies are correlated with the regressors. A form of the Chamberlain-Mundlak device is used to relate the logarithm of the effects to the regressors resulting in a lognormal distribution for the effects. The function describing the technology is modeled nonparametrically using penalized splines. Both Bayesian and non-Bayesian approaches to estimation are considered, with an emphasis on Bayesian estimation. A Monte Carlo experiment is used to investigate the consequences of ignoring correlation between the effects and the regressors, and choosing the wrong functional form for the technology.

Keywords: Technical efficiency; endogeneity; penalized splines; Gibbs sampling; maximum simulated likelihood; lognormal distribution

1. INTRODUCTION

The literature on the stochastic frontier model and its role in the estimation of firm efficiency is immense. An appreciation of the enormous number of developments and advances in model specification and estimation can be obtained from the review by Kumbhakar, Parmeter, and Zelenyuk (2017). Their review of 196 references describes, but is not limited to, alternative models for

Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling
Advances in Econometrics, Volume 40B, 1–28

ISSN: 0731-9053/doi:10.1108/S0731-90532019000040B002

the distributions of the inefficiency and idiosyncratic errors, handling of endogeneity, and nonparametric estimation of the production frontier. We add to this literature by considering a model which includes, simultaneously, nonparametric estimation of the frontier, and endogeneity in the form of correlation between the regressors and the inefficiency error. In contrast to previous studies that used kernel-based regression methods – see, for example, Fan, Li, and Weersink (1996), Parmeter and Racine (2012), and Noh (2014) – we use splines to model nonparametric elements in the frontier. To model the correlation between the regressors and the inefficiency error, we follow Griffiths and Hajargasht (2016) who used a Chamberlain-Mundlak device – see Mundlak (1978), Chamberlain (1984), and Wooldridge (2010). A number of other approaches for modeling endogeneity have been described by Amsler, Prokhorov, and Schmidt (2016).

In Section 2 we introduce a panel stochastic frontier model characterized by a flexible functional form for the frontier and a firm inefficiency error that follows a lognormal distribution with scale parameter that depends on time-invariant functions of the inputs. In Section 3, we show how a spline-based representation of the flexible frontier with a single input can be written as a linear function of a number of parameters. Estimation of the resulting model via a within estimator, a maximum likelihood estimator, and a Bayesian estimator is considered in Sections 4, 5, and 6, respectively. Changes necessary for a more general frontier with multiple inputs are described in Section 7. In Sections 8 and 9, a Monte Carlo experiment is used to assess the performance of the Bayesian estimator and to investigate the consequences of misspecification. Concluding remarks are offered in Section 10.

2. THE MODEL

We consider the following panel random effects stochastic production frontier model with a time-invariant inefficiency term:

$$y_{it} = f(\mathbf{x}_{it}) - u_i + v_{it} \tag{2.1}$$

In Eq. (2.1), $i = 1, \ldots, N$ indexes the firms and $t = 1, \ldots, T$ indexes time, $\mathbf{x}_{it}$ is a row vector of inputs, y_{it} represents the logarithm of output, $f(\mathbf{x}_{it})$ is the log of the production frontier, u_i is a non-negative random error which accounts for time-invariant inefficiency of firm i, and v_{it} is an idiosyncratic error assumed to be *i.i.d.* $N(0, \sigma^2)$. The model can also represent a stochastic cost frontier, with y_{it} being the logarithm of cost, by changing "$-u_i$" to "$+u_i$."

To model correlation between the inefficiency error u_i and some or all of the inputs, we assume $\ln(u_i)$ is normally distributed with a mean that depends on the firm averages of some of the inputs or functions of them. Averages of these functions of the inputs are collected in the vector $\mathbf{z}_i$. The resulting endogeneity model for describing how the inefficiency error is correlated with the inputs is given by:

$$\ln(u_i) = \mathbf{z}_i\boldsymbol{\gamma} + e_i \tag{2.2}$$

with $e_i \sim i.i.d.\ N(0, \lambda^2)$. Alternatively, we write $u_i \sim LN(\mathbf{z}_i\boldsymbol{\gamma}, \lambda^2)$ where LN denotes the lognormal distribution.[1] Eq. (2.2) is an extension of the model considered by Mundlak (1978) for a conventional random effects panel data model with correlated effects. Modeling of endogeneity in this way, and its extension by Chamberlain (1984), have been referred to as the Chamberlain-Mundlak device, a device which has proved to be very useful in the context of nonlinear panel data models with endogeneity. It has been applied to model endogeniety in probit, fractional response, Tobit, sample selection, count data, double hurdle, unbalanced panel models, and models with cluster sampling. See Wooldridge (2010) for a review and for references to these applications. Griffiths and Hajargasht (2016) employ this model in the context of a frontier than can be written as a linear function of parameters $f(\mathbf{x}_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}$. In this chapter, we allow for a more flexible functional form by introducing spline-based nonparametric modeling of the frontier $f(\mathbf{x}_{it})$, in the context of Eqs. (2.1) and (2.2).

3. SPECIFICATION FOR A UNIVARIATE REGRESSOR

There are a number of approaches that can be used to model the function $f(\mathbf{x}_{it})$. Studies that have done so nonparametrically, but in the context of kernel estimation, and without allowing for endogeneity, include Fan et al. (1996), Kumbhakar, Park, Simar, and Tsionas (2007), Martins-Filho and Yao (2015), Parmeter and Racine (2012), and Noh (2014). Reviews of nonparametric contributions have been provided by Parmeter and Kumbhakar (2014) and Kumbhakar et al. (2017). Here, we use penalized low-ranked splines to model $f(\mathbf{x}_{it})$. Because they have a nice Bayesian counterpart, they can be used effectively to estimate our model, and they have good properties (see e.g., Chib, Greenberg, & Jeliazkov, 2009; Claeskens, Krivobokova, & Opsomer, 2008; Ruppert, Wand, & Carroll, 2003). One can replace the nonparametric function with terms that are linear in the parameters and place Gaussian priors on these parameters.

Given the inherent difficulties of nonparametric modeling in many dimensions, and for expositional convenience, we begin by explaining the procedure for a univariate regressor x_{it}, in the context of the univariate model:

$$y_{it} = f(x_{it}) - u_i + v_{it} \tag{3.1}$$

Extension to various multivariate settings such as partially linear, additive, additive with interactions, and fully nonparametric models is straightforward and presented later. The polynomial spline representation of the function $f(x_{it})$ is[2]

$$f(x_{it}) = \beta_0 + \beta_1 x_{it} + \cdots + \beta_p {x_{it}}^p + \sum_{k=1}^{K} w_k (x_{it} - \kappa_k)_+^p \tag{3.2}$$

where $\{\kappa_1, \kappa_2, \ldots, \kappa_K\}$ are some chosen points in the range of x known as knots, and $(x_{it} - \kappa_k)_+^p$ is a polynomial term equal to zero when $x_{it} < \kappa_k$; the β_j and w_k

are coefficients to be estimated.[3] A polynomial degree of $p = 2$ is typically adequate – one less than the polynomial degree is the degree of differentiability at all points. To avoid overfitting, a restriction is imposed on the magnitude of the elements of $\mathbf{w}' = (w_1, w_2, ..., w_K)$, either of the form $\mathbf{w}'\mathbf{w} \leq C$, leading to penalized least squares or penalized likelihood estimators, or by assuming $\mathbf{w} \sim N(\mathbf{0}, \tau^2\mathbf{I})$, leading to Bayesian or non-Bayesian mixed model estimators.[4] Combining Eqs. (3.1) with (3.2), and writing the result in matrix notation convenient for estimation, we have:

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w} - \mathbf{u} \otimes \mathbf{i}_T + \mathbf{v} = X\boldsymbol{\beta} - \mathbf{u} \otimes \mathbf{i}_T + \mathbf{v} \tag{3.3}$$

where $\mathbf{y}$ and $\mathbf{v}$ are NT-dimensional vectors containing the y_{it} and v_{it}, respectively; $\mathbf{X} = [\mathbf{X}_0, \ \mathbf{X}_1]$ is an $[NT \times (p + 1 + K)]$ matrix with $\mathbf{x}_{it} = (\mathbf{x}_{0,it}, \ \mathbf{x}_{1,it})$ in its it-th row, where $\mathbf{x}_{0,it} = (1, x_{it}, ..., x_{it}^p)$, and $\mathbf{x}_{1,it} = ((x_{it} - \kappa_1)_+^p, ..., (x_{it} - \kappa_K)_+^p)$. The remaining notation in Eq. (3.3) is $\mathbf{u}' = (u_1, u_2, ..., u_N)$; $\mathbf{i}_T$ is a T-dimensional vector of ones, and $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_0, \mathbf{w}') = (\beta_0, ..., \beta_p, w_1, ..., w_K)$. Here and in what follows we use the notation $\mathbf{x}_{it}$ to denote the regressors required for the spline formulation rather than a vector of logs of inputs as was the case in Eq. (2.1). When more than one regressor is introduced, the nonparametric function f is again replaced by $\mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w}$ with a similar penalty on $\mathbf{w}$, but, as we see below, the definitions of $\mathbf{X}_0$, $\mathbf{X}_1$, and $\mathbf{x}_{it}$ change. Thus, estimation using Eq. (3.3) is equally applicable to the case of multivariate regressors, providing suitable changes are made for $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{x}_{it}$. We consider three methods of estimation, a within estimator, a maximum likelihood estimator, and a Bayesian estimator.

4. A WITHIN ESTIMATOR

One way of estimating a stochastic frontier model with correlated effects is to appeal to a fixed effects framework. Such an approach has been studied by Park, Sickles, and Simar (1998), Adams, Berger, and Sickles (1999), and Sickles (2005). These studies consider a parametric technology or use a kernel smoothing approach to model nonparametric elements. In this section we briefly discuss how the fixed effects approach can be combined with low-ranked splines to estimate the proposed model. We do not study this approach in full detail; our purpose is to demonstrate that the penalized splines framework is straightforward and appealing even if a fixed effects approach is desired. Because the within transformation eliminates the inefficiency error, the assumption in Eq. (2.2), describing the nature of the endogeneity, has no bearing on the estimation procedure.

To define the "within estimator" we average Eq. (3.1) over t to obtain:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} f(x_{it}) - u_i + \bar{v}_i \tag{4.1}$$

Subtracting Eq. (4.1) from Eq. (3.1) gives:

$$y_{it} - \bar{y}_i = f(x_{it}) - \frac{1}{T}\sum_{t=1}^{T} f(x_{it}) + v_{it} - \bar{v}_i. \tag{4.2}$$

Using splines and the definitions in Eqs. (3.2) and (3.3), we can write Eq. (4.2) in the following regression spline form:

$$\mathbf{y} - \overline{\mathbf{y}} = (\mathbf{X} - \overline{\mathbf{X}})\boldsymbol{\beta} + \mathbf{v} - \overline{\mathbf{v}} \tag{4.3}$$

where $\overline{\mathbf{y}} = (\overline{y}_1, \overline{y}_2, ..., \overline{y}_N)' \otimes \mathbf{i}_T$, $\overline{\mathbf{v}} = (\overline{v}_1, \overline{v}_2, ..., \overline{v}_N)' \otimes \mathbf{i}_T$, and $\overline{\mathbf{X}} = (\overline{\mathbf{x}}_1', \overline{\mathbf{x}}_2', ..., \overline{\mathbf{x}}_N')' \otimes \mathbf{i}_T$, with $\overline{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$. Estimating this equation by least squares subject to the restriction $\mathbf{w}'\mathbf{w} \leq C$ leads to the estimator:

$$\hat{\boldsymbol{\beta}}_w = [(\mathbf{X} - \overline{\mathbf{X}})'(\mathbf{X} - \overline{\mathbf{X}}) + \theta\mathbf{K}]^{-1}(\mathbf{X} - \overline{\mathbf{X}})'(\mathbf{y} - \overline{\mathbf{y}}) \tag{4.4}$$

where

$$\mathbf{K} = \begin{bmatrix} \mathbf{0}_{p+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \end{bmatrix}$$

The parameter θ is the Lagrange multiplier from the restriction $\mathbf{w}'\mathbf{w} \leq C$, and serves as a smoothing parameter. Its optimal value can be obtained using a secondary optimization procedure such as cross-validation. The corresponding within estimator for the function at each data point is $\hat{\mathbf{f}}_w = \mathbf{X}\hat{\boldsymbol{\beta}}_w$. If interest centers on the inefficiency errors, one can follow Park et al. (1998) and define:

$$\begin{aligned} \hat{\alpha}_i &= \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \hat{f}_w(x_{it})\right) \\ \hat{u}_i &= \max(\hat{\alpha}_i) - \hat{\alpha}_i \end{aligned} \tag{4.5}$$

Estimation of the model using this approach is easy and has the advantage of not making any specific assumptions about the form of correlation between the effects and the regressors. However, it has important disadvantages: the estimator for u_i is consistent with respect to T not N (see Park et al. 1998) and in general the $\hat{u}_i$'s are influenced by outliers.

5. A MAXIMUM LIKELIHOOD APPROACH

If we recognize the correlated random effects nature of the inefficiency errors, an alternative to fixed effects estimation is to embed the penalized spline framework within a maximum likelihood approach. Using $\mathbf{w} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ as the penalization criterion, it is convenient to rewrite the model in Eq. (3.3) as:

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 - \mathbf{u} \otimes \mathbf{i}_T + \boldsymbol{\varepsilon} \tag{5.1}$$

where $\boldsymbol{\varepsilon} = \mathbf{X}_1\mathbf{w} + \mathbf{v}$. We assume $\mathbf{v} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ which implies $\mathbf{V} = \text{cov}(\boldsymbol{\varepsilon}) = (\sigma^2\mathbf{I} + \tau^2\mathbf{X}_1\mathbf{X}_1')$. Then, we can write the conditional likelihood function as:

$$p(\mathbf{y}|\mathbf{u}) = (2\pi)^{-nT/2}|\mathbf{V}|^{-1/2}\exp\{-(\mathbf{y} - \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{u} \otimes \mathbf{i}_T)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{u} \otimes \mathbf{i}_T)/2\} \tag{5.2}$$

The unconditional likelihood is obtained by integrating over $\mathbf{u}$:

$$p(\mathbf{y}) = \int_{\mathbf{u}} \Big[(2\pi)^{-nT/2}|\mathbf{V}|^{-1/2} \times \exp\big\{-(\mathbf{y}-\mathbf{X}_0\boldsymbol{\beta}_0+\mathbf{u}\otimes\mathbf{i}_T)'\mathbf{V}^{-1}\big(\mathbf{y}-\mathbf{X}_0\boldsymbol{\beta}_0+\mathbf{u}\otimes\mathbf{i}_T\big)/2\big\}\Big] f(\mathbf{u})d\mathbf{u}. \quad (5.3)$$

In general, this integral is intractable and direct maximization is difficult, if not impossible. A more feasible approach is to maximize a simulated likelihood function (see e.g. Train (2003) for general discussions of maximum simulated likelihood and Greene (2003) for its application to a stochastic frontier model). By averaging the function in brackets in Eq. (5.3) over a sufficiently large number of draws from the distribution of $\mathbf{u}$, we can obtain an accurate estimate of the integral Eq. (5.3), which can then be used for maximization. Assuming $u_i \sim LN\big(\mathbf{z}_i\boldsymbol{\gamma}, \lambda^2\big)$, and using an inverse cumulative distribution function (CDF) method, the i-th element in a required N-dimensional vector of draws $\mathbf{u}^{(r)}$ is given by:

$$u_i^{(r)} = \exp\Big\{\mathbf{z}_i\boldsymbol{\gamma} + \lambda\Phi^{-1}\Big(\xi_i^{(r)}\Big)\Big\} \quad (5.4)$$

where $\Phi(\cdot)$ is the standard normal CDF, and the $\xi_i^{(r)}$ are independent draws from a uniform (0,1) distribution. Using this information, the simulated log-likelihood for R draws of the N-dimensional vector $\boldsymbol{\xi}^{(r)}$ is:

$$\ln L = \ln\left\{\frac{1}{R}\sum_{r=1}^{R}(2\pi)^{-nT/2}|\mathbf{V}|^{-1/2} \times \exp\Big\{-\big[\mathbf{h}^{(r)}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma},\lambda,\boldsymbol{\xi}^{(r)}\big)\big]'\mathbf{V}^{-1}\big[\mathbf{h}^{(r)}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma},\lambda,\boldsymbol{\xi}^{(r)}\big)\big]/2\Big\}\right\} \quad (5.5)$$

where $\mathbf{h}^{(r)}\big(\boldsymbol{\beta}_0,\boldsymbol{\gamma},\lambda,\boldsymbol{\xi}^{(r)}\big) = \mathbf{y} - \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{u}^{(r)}\otimes\mathbf{i}_T$. Eq. (5.5) is maximized with respect to the parameters $\big(\boldsymbol{\beta}_0, \sigma^2, \boldsymbol{\gamma}, \lambda^2, \tau^2\big)$. In practice, Halton draws are often used instead of uniform draws (see Greene (2003) and references cited therein for further information on Halton draws and their use in maximum simulated likelihood estimation). Once maximum likelihood estimates $\Big(\hat{\boldsymbol{\beta}}_0, \hat{\sigma}^2, \hat{\boldsymbol{\gamma}}, \hat{\lambda}^2, \hat{\tau}^2\Big)$ have been obtained, we can find a corresponding estimate of the conditional density for $\mathbf{y}$ that we denote by $\hat{p}(\mathbf{y}|\mathbf{u})$. Then, to obtain estimates for each of the u_i, we recognize that:

$$E(u_i|\mathbf{y}) = \int_0^{\infty} u_i p(\mathbf{u}|\mathbf{y})d\mathbf{u} = \frac{\int_0^{\infty} u_i p(\mathbf{y}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}{p(\mathbf{y})} \quad (5.6)$$

and estimate this mean using

$$\hat{u}_i = \frac{\sum_{r=1}^{R}\hat{u}_i^{(r)}\hat{p}(\mathbf{y}|\hat{\mathbf{u}}^{(r)})}{\sum_{r=1}^{R}\hat{p}\Big(\mathbf{y}|\hat{\mathbf{u}}^{(r)}\Big)} \quad (5.7)$$

where the elements of $\hat{\mathbf{u}}^{(r)}$ are given by $\hat{u}_i^{(r)} = \exp\left\{\mathbf{z}_i\hat{\boldsymbol{\gamma}}+\hat{\lambda}\Phi^{-1}\left(\xi_i^{(r)}\right)\right\}$. Finally, to estimate $\mathbf{w}$, we recognize that its best linear unbiased predictor, with parameters replaced by their estimates, and with $\hat{\mathbf{u}}'=(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_N)$, is:

$$\hat{\mathbf{w}} = \mathbf{X}_1'\left(\left(\hat{\sigma}^2/\hat{\tau}^2\right)\mathbf{I} + \mathbf{X}_1\mathbf{X}_1'\right)^{-1}\left(\mathbf{y} - \mathbf{X}_0\hat{\boldsymbol{\beta}}_0+\hat{\mathbf{u}}\otimes\mathbf{i}_T\right) \tag{5.8}$$

6. BAYESIAN ESTIMATION

Bayesian estimation has many appealing characteristics. It is convenient computationally, its marginal posterior densities do not condition on estimates of nuisance parameters, and prior information on firm efficiencies can improve estimator efficiency. We begin by specifying prior distributions, and then present the conditional posterior densities that can be used for Gibbs sampling.

6.1. Prior Distributions

In the context of Eqs. (3.3) and (2.2), prior distributions are needed for $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_0, \mathbf{w}')$, $\boldsymbol{\gamma}$, σ^2, and λ^2. We specify an uninformative prior for $\boldsymbol{\beta}_0$, and we constrain the magnitude of the w_k using a prior in line with the penalized likelihood approach. That is,

$$p(\boldsymbol{\beta}_0) \propto 1 \qquad \left(\mathbf{w}|\tau^{-2}\right) \sim N\left(\mathbf{0}, \tau^2\mathbf{I}\right). \tag{6.1}$$

A further (hierarchical) prior is placed on τ^2. Specifically, $\tau^{-2} \sim G(A_1, B_1)$ where $G(A_1, B_1)$ denotes a gamma density with shape and scale parameters A_1 and B_1, respectively. Its density is $p\left(\tau^{-2}\right) = \left[B_1^{A_1}/\Gamma(A_1)\right]\left(\tau^{-2}\right)^{A_1-1}\exp\left\{-B_1\tau^{-2}\right\}$. Relatively noninformative values for the hyperparameters, such as $A_1 = B_1 = 0.01$, are suitable. Similarly, for the variance of v_{it}, we use $\sigma^{-2} \sim G(A_\sigma, B_\sigma)$ and suggest $A_\sigma = B_\sigma = 0.01$.

For the inefficiency errors, where $u_i \sim LN\left(\mathbf{z}_i\boldsymbol{\gamma}, \lambda^2\right)$, and the $\mathbf{z}_i$ are time-invariant functions of x, we experimented with several alternative priors for $\boldsymbol{\gamma}$ and λ, considering in each case their implications for (1) Markov Chain Monte Carlo (MCMC) convergence and (2) the marginal prior distributions of the inefficiency errors, and their corresponding efficiencies, defined as $r_i = \exp(-u_i)$. Using these two criteria, we settled on a truncated normal prior for $\boldsymbol{\gamma}$ and two alternative priors for λ: a gamma prior on λ^{-2} and a truncated uniform prior on λ. Truncating a normal prior for $\boldsymbol{\gamma}$ to values that lead to reasonable efficiency values led to more precise estimates and improved MCMC convergence. A gamma prior for λ^{-2} is in line with most traditional priors specified for variance parameters, while use of a uniform prior for standard deviations in hierarchical models (which bear some similarity to our model) has been advocated by Gelman (2006). The notation we adopt for these priors is:

$$G \text{ prior}: \quad \boldsymbol{\gamma} \sim TN\left(\underline{\boldsymbol{\gamma}}, \mathbf{V}_\gamma; \mathbf{L}, \mathbf{U}\right) \quad \text{and} \quad \lambda^{-2} \sim G(A_\lambda, B_\lambda) \tag{6.2}$$

and

$$U \text{ prior}: \quad \boldsymbol{\gamma} \sim TN\left(\underline{\boldsymbol{\gamma}}, \mathbf{V}_{\gamma}; \mathbf{L}, \mathbf{U}\right) \quad \text{and} \quad \lambda \sim U(a_{\lambda}, b_{\lambda}). \tag{6.3}$$

The truncated normal parameters $\underline{\boldsymbol{\gamma}}$ and $\mathbf{V}_{\gamma}$ are what would be the prior mean vector and covariance matrix for $\boldsymbol{\gamma}$ if there was no truncation; $\mathbf{L}$ and $\mathbf{U}$ are vectors containing the lower and upper truncation points for each of the elements in $\boldsymbol{\gamma}$. For the uniform distribution in Eq. (6.3), a_{λ} and b_{λ} are the minimum and maximum values for λ, respectively.

Given the complexity of the model, the potential for difficulties with MCMC convergence, and the implications for the distribution of firm efficiencies, it is useful to provide guidelines for settings of the prior parameters in Eqs. (6.2) and (6.3). Previous work that assumed u_i is exponential with a constant scale parameter often used a relatively noninformative prior for that parameter such that the median of the resulting prior for efficiency, $r_i = \exp(-u_i)$, is 0.87. See, for example, Koop and Steel (2001). Following this lead, a good starting point for our model, where u_i follows a lognormal distribution, is to examine the implications of prior parameter settings for prior median efficiency. Suppose, in the first instance, that $u_i \sim LN\left(\gamma_0, \lambda^2\right)$; there are no variables $\mathbf{z}_i$ in the mean function for $\ln(u_i)$. We then need settings for $\left(\underline{\gamma}_0, V_{\gamma 0}, L_0, U_0\right)$ and either $(A_{\lambda}, B_{\lambda})$ or $(a_{\lambda}, b_{\lambda})$. The median of u_i is $\exp(\gamma_0)$ and the median of r_i is $\exp\{-\exp(\gamma_0)\}$. Thus, a value $\underline{\gamma}_0$ that leads to an efficiency distribution centered around r^* is $\underline{\gamma}_0 = \ln(-\ln(r^*))$. If we choose $r^* = 0.87$, then $\underline{\gamma}_0 = -2$ is a suitable value. Values for L_0 and U_0 can be chosen in a similar way. For example, setting $L_0 = -4$ leads to a maximum possible value for median efficiency of 0.982, and setting $U_0 = 0$ leads to a minimum possible value for median efficiency of 0.368. The value for $V_{\gamma 0}$ controls the possible spread of values for γ_0 within the truncation points. For example, $V_{\gamma 0} = 4$ implies $U_0 = 0$ and $L_0 = -4$ would each be one standard deviation from $\underline{\gamma}_0$ if the distribution was not truncated.

The prior for λ introduces extra prior uncertainty about the distribution of u_i and controls its skewness and variance. For the prior $\lambda^{-2} \sim G(A_{\lambda}, B_{\lambda})$, experimentation suggested that $A_{\lambda} = B_{\lambda} = 0.25$ is relatively noninformative, but sufficiently informative to facilitate MCMC convergence. For the prior $\lambda \sim U(a_{\lambda}, b_{\lambda})$, we recommend $a_{\lambda} = 0.1$ and $b_{\lambda} = 2$, values that were used in our Monte Carlo experiment in Section 8. To check whether the settings for $(L_0, U_0, a_{\lambda}, b_{\lambda})$ allow for a sufficiently wide range of possible efficiencies, the efficiencies corresponding to the mean values of u_i at the largest and smallest values for $(\gamma_0,\ \lambda)$ can be considered. At the upper truncation points, we find $E\left(u_i|\gamma_0 = 0, \lambda^2 = 4\right) = 7.4$, with corresponding efficiency value of $r = \exp(-7.4) = 0.0006$. The lower truncation points lead to $E\left(u_i|\gamma_0 = -4, \lambda^2 = 0.01\right) = 0.0185$, which has a corresponding efficiency value of 0.982. Thus, these prior settings accommodate a wide range of efficiency distributions.