Springer Proceedings in Business and Economics

William H. Greene Lynda Khalaf Paul Makdissi Robin C. Sickles Michael Veall Marcel-Cristian Voia *Editors*

Productivity and Inequality



Springer Proceedings in Business and Economics

More information about this series at http://www.springer.com/series/11960

William H. Greene • Lynda Khalaf • Paul Makdissi Robin C. Sickles • Michael Veall Marcel-Cristian Voia Editors

Productivity and Inequality



Editors William H. Greene Department of Economics New York University New York, NY, USA

Paul Makdissi University of Ottawa Ottawa, ON, Canada

Michael Veall Department of Economics McMaster University Hamilton, ON, Canada Lynda Khalaf Department of Economics Carleton University Ottawa, ON, Canada

Robin C. Sickles Department of Economics Rice University Houston, TX, USA

Marcel-Cristian Voia Department of Economics Carleton University Ottawa, ON, Canada

 ISSN 2198-7246
 ISSN 2198-7254
 (electronic)

 Springer Proceedings in Business and Economics
 ISBN 978-3-319-68677-6
 ISBN 978-3-319-68678-3
 (eBook)

 https://doi.org/10.1007/978-3-319-68678-3
 ISBN 978-3-319-68678-3
 (eBook)

Library of Congress Control Number: 2017963543

© Springer International Publishing AG 2018, corrected publication 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature The registered company is Springer International Publishing AG The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

This book is the result of a selection of papers accepted at the North American Productivity Workshop (NAPW), which is a major biennial conference of researchers and practitioners on productivity and efficiency issues that is held every two years. NAPW IX was being held between June 15 and June 18, 2015, in Quebec City, and the theme of the conference was productivity and inequality. The 2016 conference included several keynote speeches by leading international productivity experts and presentations of specialized productivity-related research such as Susanto Basu, Professor of Economics (Boston College); Allan Collard-Wexler, Associate Professor of Economics (Duke University); Russell Davidson, Professor of Economics (University); Erwin Diewert, Professor in the Vancouver School of Economics (University of British Columbia); Jonathan David Ostry, Deputy Director of the Research Department (IMF); Ariel Pakes, Thomas Professor of Economics (Harvard University); and Robin C. Sickles, Reginald Henry Hargrove Professor of Economics, Professor of Statistics (Rice University).

Each plenary session allowed for the presentation of recent and ongoing research in the areas of productivity, inequality, efficiency, data envelopment analysis, and index number theory. The plenary sessions were organized to cover a broad range of productivity, inequality, and efficiency topics. Basu looked at the general topic of productivity and the welfare of nations; Collard-Wexler showed how to estimate production functions with measurement error in inputs; Diewert presented a decomposition of US business sector TFP growth into technical progress and inefficiency components; Davidson presented the challenges of making statistical inference with income distributions; Ostry discussed redistribution, inequality, and growth; Pakes looked at new entries and new markets; and Sickles discussed the sources of income inequality by exploring the role of productivity growth.

The conference included many other researchers (150) from around the world (28 countries: Australia, Austria, Belgium, Brazil, Canada, China, Denmark, France, Germany, Greece, Hungary, Italy, Japan, Luxembourg, Macedonia, the Netherlands, Norway, Poland, Portugal, Romania, Russia, Spain, Sweden, Switzerland, Taiwan, Tunisia, the United Kingdom, and the United States), and the exposure of their research was a real gain for the Canadian community that works on issues of

productivity and inequality. The conference was hosted by the Department of Economics of Carleton University and organized in collaboration with Industry Canada, Network to Study Productivity in Canada from a Firm-Level Perspective, the Centre for Monetary and Financial Economics, and Bank of Canada.

The quality of the research papers presented at this conference attracted also representatives from Industry Canada, Bank of Canada, and Statistics Canada. Their participation was very important as the research ideas presented at the conference are very helpful for policy makers and can be used to formulate policies that can improve productivity performance of Canadian companies and, at the same time, can reduce inequality among Canadians.

Contents

Estimating Efficiency in the Presence of Extreme Outliers: A Logistic-Half Normal Stochastic Frontier Model with Application to Highway Maintenance Costs in England	1
Alternative User Costs, Productivity and Inequality in US Business Sectors W. Erwin Diewert and Kevin J. Fox	21
On the Allocation of Productivity Growth and the Determinants of U. S. Income Inequality Shasha Liu, Robin C. Sickles, and Shiyi Zhang	71
Frontier Estimation of a Cost Function System Model with Local Least Squares: An Application to Dutch Secondary Education Jos L. T. Blank	103
Aggregate Productivity and Productivity of the Aggregate:Connecting the Bottom-Up and Top-Down ApproachesBert M. Balk	119
Confidence Sets for Inequality Measures: Fieller-Type Methods Jean-Marie Dufour, Emmanuel Flachaire, Lynda Khalaf, and Abdallah Zalghout	143
Poverty-Dominant Marginal Transfer Reforms in Socially Risky Situations Paul Makdissi and Quentin Wodon	157
Exploring the Covariance Term in the Olley-Pakes Productivity Decomposition	169

Contents

The Decline of Manufacturing in Canada: Resource Curse, Productivity Malaise or Natural Evolution? Robert Petrunia and Livio Di Matteo	183
Flexible Functional Forms and Curvature Conditions: Parametric Productivity Estimation in Canadian and U.S. Manufacturing Industries	203
Jakir Hussain and Jean-Thomas Bernard Productivity Growth, Poverty Reduction and Income Inequality: New Empirical Evidence Mahamat Hamit-Haggar and Malick Souare	229
The Contribution of Productivity and Price Change to Farm-level Profitability: A Dual Approach Analysis of Crop Production in Norway Habtamu Alem	255
Estimation of Health Care Demand and its Implication on Income Effects of Individuals Hossein Kavand and Marcel Voia	275
Quantile DEA: Estimating qDEA-alpha Efficiency Estimates with Conventional Linear Programming Joseph A. Atwood and Saleem Shaik	305
Erratum	E1
Index	327

Estimating Efficiency in the Presence of Extreme Outliers: A Logistic-Half Normal Stochastic Frontier Model with Application to Highway Maintenance Costs in England



Alexander D. Stead, Phill Wheat, and William H. Greene

Abstract In Stochastic Frontier Analysis the presence of outliers in the data, which can often be safely ignored in other forms of linear modelling, has potentially serious consequences in that it may lead to implausibly large variation in efficiency predictions when based on the conditional mean. This motivates the development of alternative stochastic frontier specifications which are appropriate when the two-sided error has heavy tails. Several existing proposals to this effect have proceeded by specifying thick tailed distributions for both error components in order to arrive at a closed form log-likelihood. In contrast, we use simulation-based methods to pair the canonical inefficiency distributions (in this example half-normal) with a logistically distributed noise term. We apply this model to estimate cost frontiers for highways authorities in England, and compare results obtained from the conventional normal-half normal stochastic frontier model. We show that the conditional mean yields less extreme inefficiency predictions for large residuals relative to the use of the normal distribution for noise.

Keywords Stochastic frontier · Normal · Logistic · Outliers · Maximum simulated likelhood

W.H. Greene (⊠) Department of Economics, Stern School of Business, New York University, New York City, NY, USA e-mail: wgreene@stern.nyu.edu

© Springer International Publishing AG 2018

The original version of this chapter was revised. An erratum to this chapter can be found at https://doi.org/10.1007/978-3-319-68678-3_15

A.D. Stead • P. Wheat Institute for Transport Studies, University of Leeds, Leeds, UK e-mail: a.d.stead@leeds.ac.uk

W.H. Greene et al. (eds.), *Productivity and Inequality*, Springer Proceedings in Business and Economics, https://doi.org/10.1007/978-3-319-68678-3_1

1 Introduction

The aim of frontier analysis is to estimate a frontier function based on efficient, or at least best-practice in sample, production and cost relationships against which the efficiency of firms and other decision making units (DMU) can be measured. A challenge for such analyses is dealing with the existence of noise, resulting from random shocks and measurement error in the dependent variable — in the data. In particular, in the presence of outliers, there can be a disproportionate impact on the estimated frontier and on all predictions of efficiency relative to it. The Data Envelopment Analysis (DEA) model (Charnes et al. 1978) and related mathematical programming approaches are deterministic, in that any noise present is attributed wholly to variation in efficiency, and are therefore particularly sensitive. This is also the case with some of the cruder econometric methods, such as Corrected Ordinary Least Squares (COLS). Here we focus instead on Stochastic Frontier Analysis (SFA) which should be more robust to noise given this is considered explicitly alongside inefficiency in the model formulation.

The specific motivation for this paper comes from an issue arising from the authors' work studying cost efficiency in a number of datasets. The example used in this paper is cost analysis of highways maintenance operations of local government authorities in England, which utilises bespoke data on operating and capital expenditure provided by each authority. When we compute the standard Jondrow et al. (1982) predictor, an implausibly wide range of efficiency scores is found. This issue is caused by large estimated error variances; in particular, a large VAR(u) will lead to a large spread of efficiency predictions toward the unconditional mean (Wang and Schmidt 2009). Large error variances are in our dataset caused by the presence of a relatively large number of outliers in the data, due to a combination of under- or over-reporting, unobserved investment cycle effects, and extreme weather events.

In this paper we consider methods to better deal with noise data in the stochastic frontier setting. We consider alternative methods which are better suited to handling outliers in the data, i.e. heavier tails in the error. After consideration of possible existing approaches, this leads us to propose a new stochastic frontier model with a logistic distribution for the noise error. This model is easy to estimate and has been programmed into a bespoke version of LIMDEP.

The structure of this paper is as follows: Section 2 reviews the received methods available to handle a large number of outliers in frontier analysis, and reviews the relevant literature and Sect. 3 introduces a logistic-half normal stochastic frontier (SF) models for dealing with heavy-tailed noise. Section 4 applies these models to our data on highways maintenance costs in England and compares the results to those obtained from the standard normal-half normal SF model, and Sect. 5 gives our summary and conclusions.

2 Literature Review: Potential Approaches to Dealing with Outliers

2.1 Adopting Alternative Predictors for Inefficiency

Before considering amendments to the standard stochastic frontier model, it is natural to ask whether there are alternative efficiency predictors which yield more intuitive distributions of predictions. Given that in cross sectional models, point predictors are known to be inconsistent for the quantity of interest; namely the firm specific realisation of a random variable (Wheat et al. 2014), then several point and interval predictors could be candidates.

One candidate is the conditional mode predictor (Jondrow et al. 1982) which, for the normal-half normal model, treats all observations with positive (negative) residuals in the production (cost) frontier case as fully efficient; likewise in the normal-exponential model, all residuals past a certain threshold-i.e. the inverse of the product of the squared rate parameter from the exponential component and the standard deviation of the normal component-are predicted to be fully efficient. The conditional mode predictor therefore yields more intuitive efficiency predictions at the top relative to the conditional mean. This is because the conditional mean for all firms will always be less than one (for VAR(u) > 0) and, in the case of large VAR(v)i.e. data with many outliers, this difference is likely to be non-trivial even for the best performing DMU (due to substantial shrinkage to the unconditional mean (Wang and Schmidt 2009)). Furthermore, for all other observations the conditional mode predictor yields a predicted efficiency score higher than that from the conditional mean predictor; the latter difference, however, tends to be small in magnitude at the bottom, and its usefulness in remedying implausibly low efficiency scores is therefore limited.

Another approach is to calculate prediction intervals, which show the range of plausible efficiency predictions for a given observation. Since in the normal-half normal case the conditional distribution of u is that of a truncated normal random variable (Jondrow et al. 1982), Horrace and Schmidt (1996) propose simply using the quantile function for this distribution to compute the upper bound of a prediction interval, which is also derived by Bera and Sharma (1999). However, Wheat et al. (2014) note that this method does not necessarily yield a minimum width interval, and derive minimum width intervals for the normal-half normal case, and discuss various methods of accounting for parameter uncertainty in computing prediction intervals. The use of prediction intervals in cases where predicted efficiency values are at the extremes could be useful in that they allow us to qualify our point predictions of efficiency by explicitly recognising that there are in fact a range of probable values which efficiency can take; however, this is not a solution to the underlying problem and of course, the range of probable values will include values even more implausible than the point predictor.

Overall, while alternative predictors are useful in SFA in general, the mass of the conditional distribution for the most efficient firm in our sample is still far from

zero (even if the peak of the distribution—i.e. the mode—is zero. Thus the question remains as to whether an alternative formulation of the stochastic frontier model could yield a more intuitive distribution of efficiency predictions. In particular a formulation which puts more weight on outlying observations being the result of noise rather than inefficiency seems to be appropriate. We now consider possible means to achieve this.

2.2 Heteroskedastic Stochastic Frontier Models

The basic SFA model assumes that both error components are homoskedastic, i.e. that they have a constant variance. Outliers in the data could result from heteroskedasticity in one or both error components, so that certain observations have a higher error variance than others. Discussion of heteroskedastic SF models have tended to focus on heteroskedasticity in the one-sided error; Reifschneider and Stevenson (1991) propose a normal-half normal model in which $\sigma_{ui} = g(U_i)$, $g(U_i) \in (0, \infty)$, Caudill and Ford (1993) propose a normal-half normal model in which $\sigma_{ui} = \sigma_u (U_i \gamma)^{\delta}$, and Caudill et al. (1995) propose a normal-half normal in which $\sigma_{ui} = \exp(U_i \gamma)$, where in each case U_i is a vector of explanatory variables including an intercept. Wang (2002) combined the Battese and Coelli (1995) specification of the pre-truncation mean of a truncated normal one-sided error in which $\mu_i = Z_i\beta$, where Z_i is again a vector of explanatory variables, with a slight variation in the Caudill et al. (1995) specification of the one-sided error variance so that $\sigma_{ui}^2 = \exp(U_i\gamma)$ into a single model, which has the additional advantage of allowing for non-monotonic relationships between inefficiency and explanatory variables.

In terms of handling outliers where these are assumed to reflect an unusually high variance in noise, it is more useful to allow for heteroskedasticity in the two-sided error, however; Wang and Schmidt (2009) show for the normal-half normal model that $E(u_i | \varepsilon_i)$ is a shrinkage of u_i towards $E(u_i)$, and that because of this, as $\sigma_{vi} \rightarrow 0$, $E(u_i | \varepsilon_i) \rightarrow u_i$, while as $\sigma_{vi} \rightarrow \infty$, $E(u_i | \varepsilon_i) \rightarrow E(u_i)$. Allowing for heteroskedasticity in v therefore allows for varying levels of shrinkage. Hadri (1999) introduces a doubly heteroskedastic SF model in which the variances of both error components are a function of vectors of explanatory variables U_i and V_i —which need not be the same—such that $\sigma_{ui} = \exp(U_i\gamma)$, $\sigma_{vi} = \exp(V_i\theta)$. Finally, Kumbhakar and Sun (2013) introduce a normal-truncated normal model which combines the Battese and Coelli (1995) and Hadri (1999) specifications into a model in which the pre-truncation mean of the one-sided error, as well as the variances of both error components are functions of vectors of explanatory variables, so that $\mu_i = Z_i\beta$, $\sigma_{ui} = \exp(U_i\gamma)$, $\sigma_{vi} = \exp(V_i\theta)$.

Allowing for greater levels of variance in outlying observations is effectively another method of allowing for a heavy tailed distribution. The problem with adopting this approach using existing heteroskedastic SF models is that an appropriate variable is needed for inclusion in the variance function. A dummy variable identifying outlying observations could be used, for example, however the identification of such outlying observations would either have to be done on an ex-post basis, or with reference to some arbitrary partial metric, and of course there is an added degree of arbitrariness in defining the cut-off point beyond which an observation is deemed to be outlying.

2.3 Thick Frontier Analysis

Berger and Humphrey (1991, 1992) introduced Thick Frontier Analysis (TFA), which is motivated by the observation of heavy-tailed errors in cost studies specifically, in the banking sector—but in contrast to the present study assumes that this reflects a wide spread of efficiencies, rather than outliers in the data. In TFA, DMUs are sorted into quantiles based on some partial measure, e.g. unit cost, and separate regressions are run for the top and bottom quantiles. DMUs in the lowest and highest unit cost quantiles are implicitly judged to be equally efficient, with their residuals reflecting only error and luck. The difference in predicted unit costs for different size classes is then decomposed into exogenous market factors, i.e. that explained by differences in output mix, input prices, etc., and the remainder, which is regarded as inefficiency.

TFA has a number of disadvantages, such as the implicit assumption of equal efficiency among DMUs in the same quantile, and the implicit need for rather large sample sizes so that samples can be sensibly divided in this way. Also problematic is the arbitrariness of both the partial measure according to which DMUs are placed into quantiles, and the number of quantiles specified; Wagenvoort and Schure (1999) provide a solution to the latter problem, using a recursive algorithm by which, starting with OLS on the full sample of observations, the sample is divided into successively larger numbers of quantiles until the Lagrange multiplier test proposed by Breusch and Pagan (1980) fails to reject normality of the error term. However, the successive increases in the number of quantiles will require larger and larger sample sizes, and will tend to increase the distortionary effect of outlying observations on the estimated quantile regression lines, and hence on efficiency predictions.

The impact of outliers on efficiency scores in TFA is somewhat ambiguous. On one hand, the impact of outliers on efficiency scores will tend to be muted by the attribution of the residuals from the quantile regressions to noise, and by construction the DMUs in the top quantile will be judged fully efficient, while on the other hand the quantile regressions themselves will be more sensitive to outliers, which could lead to an exaggerated gap between the quartile regression lines, and hence an exaggerated range of inefficiency scores. This in fact reflects the different motivations and assumptions behind TFA, since as stated above, the underlying assumption behind TFA is that heavy tailed errors reflect a wide spread of inefficiency, i.e. a heavy tailed distribution of inefficiency, rather than a heavy tailed distribution of noise, making TFA inappropriate for the purpose of the current study; we therefore do not pursue TFA any further.

2.4 Non-Gaussian Stochastic Frontier Models

Another possible method of dealing with the impact of outliers in the data on efficiency scores is to directly alter the distributional assumptions of the basic SF model such that the noise component of the composed error, rather than being normally distributed, follows an alternative symmetric distribution with heavier tails.

One candidate for this is the Student's t distribution, a heavy-tailed distribution which approximates normality for finite sample sizes. Tancredi (2002) proposes a model in which the two-sided error is t distributed and the one-sided error follows a half t distribution—thus generalising the original normal-half normal of Aigner et al. (1977) to allow for heavier tails in both components of the composed errorand shows that as the residual approaches infinity, the conditional distribution of the one-sided error (conditional on the composed error realisation) is concentrated around zero in the normal-half normal model, and is completely flat in the *t*-half t model; thus in the former case, an observation with a large positive residual is judged to be close to the frontier with high probability, while in the latter case it is judged to be basically uninformative, making the model better at handling such outliers. Applying both models to the Christensen and Greene (1976) dataset on US electric utilities, the author shows that the *t*-half *t* performs better than the normalhalf normal, and that allowing for heavy tails in this way increases the evidence for inefficiency in the model and overturns the Ritter and Simar (1994) finding that the basic SF model does not fit the data significantly better than OLS.

Nguyen (2010) introduces three additional non-Gaussian SF models, having twosided and one-sided errors that respectively follow Laplace and exponential, Cauchy and half Cauchy, and Cauchy and truncated Cauchy distributions. These models are considered in a cross-section context, with application to the Christensen and Greene (1976) dataset, and Cauchy-half Cauchy balanced and unbalanced panel data models with time invariant inefficiency are also introduced, with application to the US banking dataset and to the WHO health sector dataset used in Greene (2004). The usefulness of some of the aforementioned models is limited by the unjustifiable assumptions made in order to simplify their derivation: the Laplace-exponential model assumes the variances of the two error components to be the same, as does the Cauchy-half Cauchy model for balanced panel data with respect to the variance of the two-sided error and the (pre-truncation) variance of the one-sided error; the latter model further assumes only two time periods. Nevertheless, both the crosssection and unbalanced panel Cauchy-half Cauchy models appear acceptable, and results from the latter are presented by Gupta and Nguyen (2010).

Horrace and Parmeter (forthcoming) discuss SFA with a Laplace-distributed two-sided error generally, and introduce a Laplace-truncated Laplace model; this is shown to reduce to a Laplace-exponential model when the pre-truncation mean of the one-sided error is less than zero, and to a Least Absolute Deviations (LAD) regression when the variance of the inefficiency term is zero. It is also shown that the conditional distribution of inefficiency is constant when the residual is zero, so that all observations with positive residuals are given an identical efficiency score; as with the t-half t, the model therefore treats outlying observations as less informative. Results from Monte Carlo simulations suggest that the Laplace-exponential model performs better than the normal-exponential model when the error is miss-specified, and that it is more likely to produce non-zero estimates of the variance in inefficiency when OLS residuals display the wrong skew. The Laplace-truncated Laplace model is applied to estimate a cost frontier using the US airline data used in Greene (2012).

An analogous Bayesian approach to non-Gaussian SFA exists; Tchumtchoua and Dey (2007), estimate a *t*-half *t* Bayesian SF model, and Griffin and Steel (2007) briefly discuss how to estimate *t*-half normal, *t*-exponential, and *t*-gamma Bayesian SFA models using the WinBUGS software package.

To summarize, the non-Gaussian SF models are a potential way of dealing with the impact of outliers on the spread of efficiency predictions in SFA, given the different way the models treat outliers; they also have the advantage of being less arbitrary than simply excluding observations, or than the other methods discussed. A drawback of the existing models, however, is that in order to arrive at closed form expressions for their log-likelihoods, they also adopted alternative—i.e. thick tailed—distributions for u, which limits both the effectiveness of the models in reducing the impact of outliers on the range of efficiency predictions, and comparability with conventional SF models; we therefore prefer a model in which only v is drawn from a thick tailed distribution.

3 The Logistic-Half Normal Stochastic Frontier Model

3.1 Formulation and Estimation

In this paper, our motivation is to amend the conventional stochastic frontier model to accommodate data with large reporting errors. The work on non-Gaussian SF models discussed above motivates us to propose a further model which departs from the previous literature in that it amends the noise error term only and retains all of the conventional SF assumptions on the inefficiency error and the relationship between error components and regressors. This allows us to understand the extent to which alternative assumptions on the noise error term influence the efficiency predictions all other things equal.

In SFA, we have a composed error ε consisting of a symmetric noise component v and an inefficiency component u which is drawn from some one-sided distribution, such that

$$\varepsilon = v - su \tag{1}$$

Where *s* takes on a value of one for a production frontier and minus one for a cost frontier. In our case, we assume that v is drawn from a logistic distribution, and that v is from a half-normal distribution, such that

$$f(v) = \frac{\exp\left(\frac{v}{\sigma_v}\right)}{\sigma_v \left[1 + \exp\left(\frac{v}{\sigma_v}\right)\right]^2}$$
(2)

$$f(u) = \begin{cases} \frac{2}{\sigma_u} \phi\left(\frac{u}{\sigma_u}\right), & su > 0\\ 0, & su \le 0 \end{cases}$$
(3)

Where σ_v and σ_u are scale parameters. The joint density of ε and u is given by

$$f(u,\varepsilon) = \begin{cases} \frac{\exp\left(\frac{\varepsilon+su}{\sigma_v}\right)}{\sigma_v \left[1+\exp\left(\frac{\varepsilon+su}{\sigma_v}\right)\right]^2} \frac{2}{\sigma_u} \phi\left(\frac{u}{\sigma_u}\right), & su > 0\\ 0, & su \le 0 \end{cases}$$
(4)

And the marginal density of ε is given by the convolution

$$f(\varepsilon) = \int_{0}^{\infty} \frac{\exp\left(\frac{\varepsilon + su}{\sigma_{v}}\right)}{\sigma_{v} \left[1 + \exp\left(\frac{\varepsilon + su}{\sigma_{v}}\right)\right]^{2}} \frac{2}{\sigma_{u}} \phi\left(\frac{u}{\sigma_{u}}\right) du$$
(5)

Which is an integral with no closed form. It is therefore not possible to give an analytic expression for the log-likelihood function, and to proceed with maximum likelihood estimation. In such a case, maximum simulated likelihood techniques—see Train (2003) for an introduction to simulation-based methods—allow us to overcome this obstacle and estimate our model. The method followed here was first outlined in the context of the normal-gamma SF model by Greene (2003). We begin by noting that the integral in (5) is simply the expectation of f(v) given that u is drawn from a half normal distribution

$$h(u) = E[f(v)|u \ge 0], \quad u \sim N[\mu, \sigma_u]$$
(6)

And thus we can form a simulated probability density function for ε by averaging over Q draws from a half normal distribution. The usual method of taking draws from a non-uniform distribution is to note that the cumulative density function of a random variable follows a uniform distribution, and thus by inverting the cumulative density function we can have the value of the random variable in terms of a uniformly distributed random variable; this inverse cumulative density function can therefore be used to transform draws from a uniform distribution into draws from any given distribution. Thus to generate draw number q from the half normal Estimating Efficiency in the Presence of Extreme Outliers...

distribution of our inefficiency term u we have

$$u_q = \sigma_u \Phi^{-1} \left(\frac{1}{2} + \frac{F_q}{2} \right) \tag{7}$$

Where F_q is draw number q from a uniform distribution. This leads us to the simulated probability density function for ε

$$\tilde{f}(\varepsilon) = \frac{1}{Q} \sum_{q=1}^{Q} \frac{\exp\left(\frac{\varepsilon + su_q}{\sigma_v}\right)}{\sigma_v \left[1 + \exp\left(\frac{\varepsilon + su_q}{\sigma_v}\right)\right]^2}$$
(8)

And, introducing subscripts for observation i, the simulated log-likelihood function is

$$\ln SL = -N \ln Q - N \ln \sigma_v + \sum_{i=1}^N \ln \sum_{q=1}^Q \frac{\exp\left(\frac{\varepsilon_i + su_{qi}}{\sigma_v}\right)}{\left[1 + \exp\left(\frac{\varepsilon_i + su_{qi}}{\sigma_v}\right)\right]^2}$$
(9)

Which may be maximised like any conventional log-likelihood function, provided we have our draws from the uniform distribution forming the u_{ai} s.

3.2 Efficiency Predictions

The conditional density of *u* given ε , is the ratio of the joint distribution of *v* and *u* and the density of ε

$$f(u|\varepsilon) = \frac{f(v)f(u)}{f(\varepsilon)}$$
(10)

Which, in the logistic-half normal case, gives

The Jondrow et al. (1982) and Battese and Coelli (1988) point predictors for efficiency are $\exp[-E(u|\varepsilon)]$ and $E[\exp(-u|\varepsilon)]$, respectively; these are derived by solving the integrals

$$E(u|\varepsilon) = \int_{0}^{\infty} uf(u|\varepsilon) du$$
(12)

$$E\left[\exp(-u)|\varepsilon\right] = \int_{0}^{\infty} \exp(-u)f\left(u|\varepsilon\right) du$$
(13)

Which, in the logistic-half normal case, gives

$$E(u|\varepsilon) = \frac{1}{f(\varepsilon)} \int_{0}^{\infty} \frac{u \exp\left(\frac{\varepsilon + su}{\sigma_{v}}\right)}{\sigma_{v} \left[1 + \exp\left(\frac{\varepsilon + su}{\sigma_{v}}\right)\right]^{2}} \frac{2}{\sigma_{u}} \phi\left(\frac{u}{\sigma_{u}}\right) du$$
(14)

$$E\left[\exp(-u)|\varepsilon\right] = \frac{1}{f(\varepsilon)} \int_{0}^{\infty} \frac{\exp(-u)\exp\left(\frac{\varepsilon+su}{\sigma_{v}}\right)}{\sigma_{v} \left[1 + \exp\left(\frac{\varepsilon+su}{\sigma_{v}}\right)\right]^{2}} \frac{2}{\sigma_{u}} \phi\left(\frac{u}{\sigma_{u}}\right) du \qquad(15)$$

Both of which, again, contain integrals with no closed form solutions. Simulation is therefore required to generate these point predictions: we substitute $\tilde{f}(\varepsilon)$ for $f(\varepsilon)$, and the remaining integrals are the expectation of u and $\exp(su)$ respectively multiplied by the probability density function of v, given that u is drawn from a half-normal distribution; this leads us to the simulated expectations

$$\tilde{E}(u|\varepsilon) = \frac{1}{\tilde{f}(\varepsilon)} \frac{1}{R} \sum_{r=1}^{R} \frac{u_r \exp\left(\frac{\varepsilon + su_r}{\sigma_v}\right)}{\sigma_v \left[1 + \exp\left(\frac{\varepsilon + su_r}{\sigma_v}\right)\right]^2}$$
(16)

$$\tilde{E}\left[\exp(-u)|\varepsilon\right] = \frac{1}{\tilde{f}(\varepsilon)} \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left[\frac{\varepsilon + (s+\sigma_v)u_r}{\sigma_v}\right]}{\sigma_v \left[1 + \exp\left(\frac{\varepsilon + su_r}{\sigma_v}\right)\right]^2}$$
(17)

Which we use to generate our point predictions of cost efficiency. Note that draws from the uniform distribution are also therefore needed to generate efficiency predictions following estimation of the model. In the notation above we distinguish between draws to approximate $f(\varepsilon)$ using q and the additional draws required to compute the further integral in (16) and (17) using r. This is to minimise any simulation bias.

4 Application to Highways Maintenance Costs in England

In this section, we apply the logistic-half normal SF model to a unique dataset on highway maintenance costs in England. Responsibility for maintaining roads in England is divided between Highways England—until 2015 the Highways Agency—a government-owned company responsible for maintenance of the trunk road network, and the county councils and unitary authorities which are responsible for maintenance of the non-trunk roads in their respective areas. In recent years, local authorities have been under increasing pressure to demonstrate efficient practice or efficiency improvements in areas such as highway maintenance, e.g. by undertaking benchmarking exercises with peers. This study uses data from the CQC Efficiency Network,¹ which is used to analyse the cost efficiency of local authorities' highway maintenance activities.

Previous econometric studies of road maintenance costs have tended to focus of the question of marginal costs of usage, and what these imply for road pricing, rather than on the relative cost efficiency of local authorities. Previous studies estimate cost functions using data on renewals and maintenance costs for motorways and canton roads in Switzerland (Schreyer et al. 2002), Austrian motorways (Sedlacek and Herry 2002), national-i.e. trunk-roads in Poland (Bak et al. 2006; Bak and Borkowski 2009), roads in Sweden (Haraldsson 2006; Jonsson and Haraldsson 2008), and German motorways (Link 2006, 2009) and federal roads (Link 2014). Much of this work is summarized by Link (2014), who estimates two cost models: one in which, as the author argues should be the case, the size of the road network maintained is used as the scale variable, and a second in which passenger car traffic and goods vehicle traffic are used as scale variables can be derived; the author apparently does not consider using both network size and traffic as outputs in a single model. The only study to look at efficiency in the context of highway maintenance is that of Fallah-Fini et al. (2009), which uses applies DEA to data for eight counties from the US state of Virginia, using road area and a set of quality measures as outputs, and maintenance expenditure, traffic and equivalent single axle loads as inputs, and a set of climate factors as non-discretionary variables.

We use an unbalanced panel consisting of data on the 70 local authorities from England that were members of the CQC efficiency network during 2014–15 and supplied cost data for at least one of the 5 years from 2009–10 to that year; this gives us a total of 327 observations. Cost data were supplied to the network by each authority individually according to definitions decided by a working group of network members, relating to operating expenditure and capital expenditure both divided into direct and indirect categories—on carriageway maintenance only, i.e. excluding related activities such as winter service and footway maintenance, on the basis that they should be understandable and yield consistent submissions; we use the sum of these, total expenditure, as our dependent variable. Nevertheless,

¹See http://www.nhtnetwork.org/cqc-efficiency-network/home/

(18)

preliminary analysis of the data reveals large differences in unit costs with a large number of extreme outliers in both direction, which are clearly subject to some kind of reporting error. As a result, standard SF models, as discussed in Sect. 1, yields a wide range of efficiency predictions, motivating the development of the model presented here.

In line with the previous literature, we use road length and traffic as output variables; road lengths are included as our measure of scale, while traffic—in terms of passenger kilometres—we divide by road length and include as a density variable. Detailed breakdowns of overall network length into urban and rural roads and also by classification, the different classifications being, in order of importance, A roads, B roads, classified unnumbered roads, and unclassified roads; we refer to the latter two as C and U roads, respectively. B, C and U roads are always maintained by local authorities, while A roads can be either trunk, and therefore the responsibility of Highways England, or non trunk, maintained by local authorities. The road length data we use include B, C, U and non trunk A roads; motorways, denoted by the letter M, and trunk A roads, are not included. Likewise, we use traffic data supplied directly by the Department for Transport (DfT) which relate only to local-authority maintained roads.

We separate overall network length into urban and rural road lengths, and further include the lengths relating to each classification as proportions of the overall network length. We also include road condition indicators for each road classification—also from DfT sources—and as input prices we include a measure of median hourly wages in civil engineering for each NUTS1 region from the Annual Survey of Hours and Earnings (ASHE) published by the Office for National Statistics (ONS) and a national index of materials prices in road construction from the Department for Business, Innovation and Skills (BIS). We employ a modified Cobb-Douglas functional form, in which we include second-order terms relating to urban and rural road length. The cost frontier we estimate is

$$\begin{aligned} \ln TOTEX &= \beta_0 + \beta_1 \ln URL + \beta_2 \ln RRL + \beta_3 \ln URL^2 + \beta_4 \ln RRL^2 \\ &+ \beta_5 \ln URL \ln RRL + \beta_6 \ln TRAFFIC + \beta_7 RDCA + \beta_8 RDCBC + \beta_9 RDCU \\ &+ \beta_{10} PROP_{UA} + \beta_{11} PROP_{UB} + \beta_{12} PROP_{UC} + \beta_{13} PROP_{UU} \\ &+ \beta_{14} PROP_{RA} + \beta_{15} PROP_{RB} + \beta_{16} PROP_{RC} + \beta_{17} YEAR + \beta_{18} \ln WAGE \\ &+ \beta_{19} \ln ROCOSM + \varepsilon \end{aligned}$$

Where *TOTEX* is total expenditure on carriageway maintenance, *URL* and *RRL* are the lengths of an authority's urban and rural road networks, respectively, *TRAFFIC* is a traffic density measure—i.e. traffic count divided by total road network length—and *RDCA*, *RDCBC* and *RDCU* are the proportions of A roads, B and C roads, and unclassified roads where maintenance should be considered, weighted by the shares of their respective road classifications in the total road network length. *PROP*_{UA} through to $PROP_{RC}$ are urban A roads, urban B roads, etc. as proportions of the total network length, with the proportion of rural unclassified

	Logistic-Half Normal			Normal-Half Normal		
	Estimate	s.e.	Sig	Estimate	s.e.	Sig
β_0	16.0631	0.0956	***	16.0350	0.14502	***
$\beta_1 (\ln URL)$	0.13443	0.11162		0.12738	0.17112	
$\beta_2 (\ln RRL)$	0.90841	0.11836	***	0.91675	0.17943	***
$\beta_3 (\ln URL^2)$	0.23534	0.04447	***	0.24091	0.06291	***
$\beta_4 (\ln RRL^2)$	0.08315	0.01057	***	0.08503	0.01586	***
$\beta_5 (\ln URL \ln RRL)$	-0.07189	0.02944	**	-0.08083	0.04421	*
$\beta_6 (\ln TRAFFIC)$	0.37956	0.10259	***	0.41532	0.15442	***
$\beta_7 (RDCA)$	0.44014	0.09675	***	0.46356	0.14373	***
$\beta_8 (RDCBC)$	-0.07142	0.02682	***	-0.07057	0.03909	*
$\beta_9 (RDCU)$	-0.00397	0.00324		-0.00519	0.00529	
$\beta_{10} (PROP_{UA})$	8.28742	1.9879	***	7.80954	3.24067	**
$\beta_{11} (PROP_{UB})$	1.982	2.27009		0.66161	3.86852	
$\beta_{12} (PROP_{UC})$	0.62504	1.21835		0.44784	2.05441	
$\beta_{13} (PROP_{UU})$	1.10074	0.56802	*	1.09028	0.83493	
$\beta_{14} (PROP_{RA})$	2.57286	1.08575	**	2.1196	1.57145	
$\beta_{15} (PROP_{RB})$	2.40330	1.10305	**	2.67772	1.5444	*
$\beta_{16} (PROP_{RC})$	1.11517	0.67064	*	0.98277	0.98812	
β_{17} (YEAR)	0.04055	0.01105	***	0.04457	0.01661	***
β_{18} (ln WAGE)	0.82267	0.23264	***	0.89086	0.34002	***
$(1-\beta_{18})$ $(\ln ROCOSM)^1$	0.17733	-	-	0.10914	-	-
σ_u	0.54321	0.02541	***	0.56798	0.01482	-
σ_v	0.16005	0.00745	***	0.27642	0.03015	-
Log Likelihood	-188.52			-189.14		

 Table 1
 Outputs from the logistic-half normal and normal-half normal models

Statistical significance at the: * 10% level, ** 5% level, *** 1% level

Notes: (1) Parameter is equivalent to $1 - \beta_{18}$ due to the imposition of linear homogeneity in input prices

roads omitted to avoid perfect multicollinearity. Finally, we include a time trend, *YEAR*, and two input prices: *WAGE*, a measure of regional gross hourly wages in civil engineering, and *ROCOSM*, a national index of materials prices for road construction. All variables are mean-centred, and linear homogeneity in input prices is imposed by dividing our cost and wage variables by our materials price index, which drops out of the model.

Table 1 shows the parameter estimates and associated standard errors and significance levels from the logistic-half normal model, and for comparison, the normal-half normal model, both estimated in LIMDEP. Following Greene (2003), we use Halton draws rather than pseudorandom number generator to obtain our draws from the uniform distribution; we use 1000 draws, and find that further increases or small reductions in the number of draws do not significantly affect our results.

We can see that both models yield similar estimates for each parameter, and that most of our variables are found to be statistically significant at the 10%, 5%, or 1% levels. To underline the similarities between the two models, we note that the correlation between the predicted residuals from each model is 0.9994 (rank correlation 0.9993). The log likelihood for the logistic-half normal model is higher than the corresponding value for the normal-half normal model indicating a superior fit.

The parameter estimates indicate constant to decreasing returns to scale at the sample average (the p-value for the null hypothesis of constant returns is 0.2396, so we fail to reject it), with increasing returns to scale for smaller authorities, and increasing returns to traffic density. It is also noticeable that the significance associated with each of the frontier parameters increases using the logistic-half normal model relative to the normal-half normal model. This is unsurprising, since the use of a thick-tailed noise distribution increases the robustness of our parameter estimates to outliers.

Also of interest here are the estimated error variances, and how these differ between the two models. The variance of u is given in both cases by

$$VAR(u) = \frac{\pi - 2}{\pi} \sigma_u^2$$
(19)

While the variances of v in the logistic-half normal and normal-half normal models, respectively, are given by

$$VAR(v) = \frac{\pi^2}{3}\sigma_v^2$$
(20)

$$VAR(v) = \sigma_v^2 \tag{21}$$

Table 2 shows VAR(u) and VAR(v) for both the logistic-half normal and normalhalf normal models, along with total error variance, VAR(ε). We can see that neither the overall error variance, nor its individual components, differ substantially between the two models.

In spite of their similar error variances, however, we expect that the logistichalf normal model will result in a significantly narrower distribution of predicted efficiency scores, given the very different way that the two models handle outliers, as discussed in Sect. 3.2. Cost efficiency predictions from both models are generated using the Jondrow et al. (1982) conditional mean predictor, which is shown in (16) for the logistic-half normal case.

Table 2 Estimated error variances		Logistic-Half Normal	Normal-half normal	
	VAR(u)	0.107225	0.117227	
	VAR(v)	0.084279	0.07641	
	$VAR(\varepsilon)$	0.191504	0.193637	

Table 3 Summary of		Logistic-Half Normal	Normal-half normal	
efficiency scores		Logistic Hair Horman	rtorinar nan normar	
	Minimum	0.408882	0.225086	
	Mean	0.708911	0.659549	
	Median	0.724585	0.682412	
	Maximum	0.879474	0.918035	
	Range	0.470592	0.692949	

Table 3 shows some summary statistics relating to the resulting efficiency predictions from both models. The correlation between the two sets of efficiency predictions is high, at 0.997. However, comparing the ranges of the two sets of predictions, we can see that, as expected, the logistic-half normal model results in a far narrower distribution of efficiency predictions. This is due mostly to a very marked difference in the minimum predicted efficiency score, which is far higher in the logistic-normal model, from which the mean and the median predictions are also higher, though the difference is progressively smaller in each case. The maximum prediction, however, is smaller in the logistic-half normal model than in the normal-half normal model due to the way the model handles outliers in either direction, though as discussed in Sect. 2.1, the maximum prediction from both models would have been one if we had used the conditional mode predictor.

Figure 1 gives a more detailed comparison, showing kernel density estimates for both sets of efficiency scores. In this, we can see a greater number of observations with low predicted efficiency scores from the normal-half normal model generally, and higher efficiency predictions generally more common in the logistic-half normal model; the latter being in spite of the fact that, due to the model's handling of outlying observations, the highest several efficiency scores are somewhat lower than those from the normal-half normal model. Our model therefore seems to result in an overall more intuitive distribution of efficiency predictions, with far fewer at the bottom of the range with only a relatively small impact on predictions at the top.

Figure 2 shows the relationship between efficiency predictions and corresponding residuals in both models. Given the similarity of the estimated frontier parameters, the ranges of the residuals across the two models are very similar, as are the estimated error variances, but the relationship between the residuals and the efficiency predictions are significantly different; in the normal-half normal model, the slope of the function diminishes for large positive or negative residuals, but in the logistic-half normal model, in addition to the slope being gentler overall, this is much more pronounced, with the function becoming almost flat — i.e. there being very little change in efficiency predictions — at either end of the range. This suggests that, in line with our discussion of the way that the model treats outlying observations, efficiency predictions do not approach zero or one for extreme values of the residuals.



Fig. 1 Kernel densities of cost efficiency scores



Fig. 2 Cost efficiency scores against residuals

5 Summary and Conclusions

This paper considers the issue of outliers and their impact on efficiency analyses. After reviewing how these issues have been handled in the existing literature, we have motivated and formulated a stochastic frontier (SF) model with a thick-tailed noise component. In contrast to previous models, in which both the noise and inefficiency terms have been drawn from a thick-tailed distribution, we use maximum simulated likelihood to estimate a model which combines a thick-tailed noise distribution—i.e. a logistic distribution—with a half normal inefficiency distribution. This model is easy to estimate and has been programmed into a bespoke version of LIMDEP. We show that the model handles outliers in both directions in a way that can produce a much narrower—and in the presence of outliers, more intuitive—range of efficiency predictions than standard SF models.

We apply our model to a unique dataset on highways maintenance costs in England, and compare the results to those from the normal-half normal SF model. The estimated frontier parameters and variances are found to be very similar to those from the normal-half normal model, but the former with greater significance due to the increased robustness of the model to outlying observations and we find, as expected, that the model results in a narrower range of efficiency predictions. The model is therefore effective in reducing the extent to which outlying observations are treated as having extreme efficiency values.

Further development could consider alternative distributions for u, such as truncated normal, exponential, or gamma, which would be easy to implement using our estimation approach. The issue of testing between our model and the standard SF model could also be explored. The authors are currently developing an alternative model in which v follows a Student's t distribution, which has the normal distribution as a limiting case, meaning that the model nests the standard SF model. A further advantage of the Student's t is that the thickness of the tails can be varied with its degrees of freedom parameter, making the model more general; a Student's t distribution used in this study.

Acknowledgements The authors acknowledge funding from the CQC Efficiency Network (see http://www.nhtnetwork.org/cqc-efficiency-network/home/).

References

- Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21–37.
- Bak, M., & Borkowski, P. (2009). Marginal cost of road maintenance and renewal in Poland, CATRIN (Cost Allocation of TRansport INfrastructure cost) deliverable D6, annex 2. Leeds: ITS, University of Leeds.
- Bak, M., Borkowski, P., Musiatowicz-Podbial, G., & Link, H. (2006). Marginal Infrastructure cost in Poland, marginal cost case studies for road and rail Transport deliverable D3, annex 1.2C. Leeds: ITS, University of Leeds.

- Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*(3), 387–399.
- Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2), 325–332.
- Bera, A. K., & Sharma, S. C. (1999). Estimating production uncertainty in stochastic frontier production function models. *Journal of Productivity Analysis*, 12(3), 187–210.
- Berger, A. N., & Humphrey, D. B. (1991). The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics*, 28(1), 117–148.
- Berger, A. N., & Humphrey, D. B. (1992). Measurement and efficiency issues in commercial banking. In Z. Griliches (Ed.), *Output Measurement in the Service Sectors. NBER* (pp. 245– 300).
- Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies.*, 47(1), 239–253.
- Caudill, S. B., & Ford, J. M. (1993). Biases in frontier estimation due to heteroscedasticity. *Economics Letters*, 41(1), 17–20.
- Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics*, 13(1), 105–111.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Christensen, L. R., & Greene, W. H. (1976). Economies of scale in U.S. electric power generation. Journal of Political Economy, 84(4), 655–676.
- Fallah-Fini, S., Triantis, K., & de la Garza, J. M. (2009). Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In *IIE* annual conference. Proceedings (p. 693). Institute of Industrial Engineers-Publisher, Miami, USA.
- Greene, W. H. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, 19(2), 179–190.
- Greene, W. H. (2004). Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics*, 13(10), 959–980.
- Greene, W. H. (2012), Econometric analysis (7th ed.) Pearson.
- Griffin, J. E., & Steel, M. F. J. (2007). Bayesian stochastic frontier analysis using WinBUGS. Journal of Productivity Analysis, 27(3), 163–176.
- Gupta, A. K., & Nguyen, N. (2010). Stochastic frontier analysis with fat-tailed error models applied to WHO health data. *International Journal of Innovative Management, Information & Production.*, *1*(1), 43–48.
- Hadri, K. (1999). Estimation of a doubly Heteroscedastic stochastic frontier cost function. *Journal of Business & Economic Statistics*, 17(3), 359–363.
- Haraldsson, M. (2006). Marginal cost for road maintenance and operation—A cost function approach, marginal cost studies for road and rail Transport deliverable D3, annex. Leeds: ITS, University of Leeds.
- Horrace, W. C., & Parmeter, C. F. (forthcoming). Forthcoming. A Laplace stochastic frontier model. *Econometric Reviews*.
- Horrace, W. C., & Schmidt, P. (1996). Confidence statements for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis*, 7(2), 257–282.
- Jondrow, J., Knox Lovell, C. A., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2), 233–238.
- Jonsson, L., & Haraldsson, M. (2008). Marginal costs of road maintenance in Sweden, CATRIN (cost allocation of TRansport INfrastructure cost) deliverable D6, annex 1. Stockholm: VTI.
- Kumbhakar, S. C., & Sun, K. (2013). Derivation of marginal effects of determinants of technical inefficiency. *Economics Letters*, 120(2), 249–253.

- Link, H. (2006). An econometric analysis of motorway renewal costs in Germany. Transportation Research Part A: Policy and Practice, 40(1), 19–34.
- Link, H. (2009). Marginal costs of road maintenance in Germany, CATRIN (Cost Allocation of TRansport INfrastructure) deliverable D6, annex 3. Stockholm: VTI.
- Link, H. (2014). A cost function approach for measuring the marginal cost of road maintenance. Journal of Transport Economics and Policy (JTEP), 48(1), 15–33.
- Nguyen, N. (2010). *Estimation of technical efficiency in stochastic frontier analysis*. PhD thesis, Bowling Green State University.
- Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review*, 32(3), 715–723.
- Ritter, C., & Simar, L. (1994). *Another look at the American electrical utility data*. CORE Discussion Paper 9407. Centre for Operations Research and Econometrics, Catholic University of Louvain.
- Schreyer, C., Schmidt, N., & Maibach, M. (2002). Road econometrics-case study motorways Switzerland. UNITE (UNIfication of accounts and marginal costs for Transport efficiency) deliverable 10, annex A1b. Leeds: ITS, University of Leeds.
- Sedlacek, N., & Herry, M. (2002). Infrastructure cost case studies. UNITE (UNIfication of accounts and marginal costs for Transport efficiency) deliverable 10, annex A1c. Leeds: ITS, University of Leeds.
- Tancredi, A. (2002). Accounting for heavy tails in stochastic frontier models. Working Paper No. 2002.16. Department of Statistical Sciences, University of Padua.
- Tchumtchoua, S., & Dey, D. K. (2007). Bayesian estimation of stochastic frontier models with multivariate skew t error terms. *Communications in Statistics - Theory and Methods.*, 36(5), 907–916.
- Train, K. E. (2003). *Discrete choice methods with simulation* (2 (2009) ed.). Cambridge University Press, Cambridge, UK
- Wagenvoort, J. L. M., & Schure, P. H. (1999). The recursive thick frontier approach to estimating efficiency. Report 99/02. European Investment Bank.
- Wang, H.-J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18(3), 241–253.
- Wang, W. S., & Schmidt, P. (2009). On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics*, 148(1), 36–45.
- Wheat, P., Greene, W., & Smith, A. (2014). Understanding prediction intervals for firm specific inefficiency scores from parametric stochastic frontier models. *Journal of Productivity Analy*sis, 42(1), 55–65.

Alternative User Costs, Productivity and Inequality in US Business Sectors

W. Erwin Diewert and Kevin J. Fox

Abstract Using the new Bureau of Economic Analysis (BEA) Integrated Macroeconomic Accounts as well as other BEA data, we construct productivity accounts for two key sectors of the US economy: the Corporate Nonfinancial Sector (Sector 1) and the Noncorporate Nonfinancial Sector (Sector 2). Calculating user costs of capital based on, alternatively, ex post and predicted asset price inflation rates, we provide alternative estimates for capital services and Total Factor Productivity growth for the two sectors. Rates of return on assets employed are also reported for both sectors. In addition, we compare rates of return on assets employed and TFP growth rates when the land and inventory components are withdrawn from the asset base. Finally, implications for labour and capital shares from using alternative income concepts are explored.

Keywords User cost of capital · Total Factor Productivity · Rate of return on assets · Integrated Macroeconomic Accounts · Bureau of Economic Analysis · Ex post and ex ante asset inflation rates · US Nonfinancial Sector · Austrian model of production · Balancing rates of return · Inequality

JEL Classification: B25, C43, C82, D24, E22, E43

© Springer International Publishing AG 2018

The first author, W. Erwin Diewert, gratefully acknowledges the financial support of the SSHRC of Canada, and both authors, W. Erwin Diewert and Kevin J. Fox, gratefully acknowledge the financial support of the Australian Research Council (LP0884095, DP150100830).

W.E. Diewert (⊠) Vancouver School of Economics, University of British Columbia, Vancouver, BC, Canada

The School of Economics, UNSW, Sydney, NSW, Australia e-mail: erwin.diewert@ubc.ca

K.J. Fox CAER & School of Economics, UNSW, Sydney, NSW, Australia e-mail: K.Fox@unsw.edu.au

W.H. Greene et al. (eds.), *Productivity and Inequality*, Springer Proceedings in Business and Economics, https://doi.org/10.1007/978-3-319-68678-3_2

1 Introduction

The US Bureau of Economic Analysis (BEA), in conjunction with the Bureau of Labor Statistics (BLS) and the Board of Governors of the Federal Reserve, have developed a new set of production accounts, the Integrated Macroeconomic Accounts, for two major private sectors of the US economy: the Corporate Non-financial Sector (which we will call Sector 1) and the Noncorporate Nonfinancial Sector (which we will call Sector 2). For both sectors we work out the rate of return on assets employed back to 1960 and compute estimates of Total Factor Productivity (TFP) growth. In addition to comparing results across the sectors, we are particularly interested in determining whether rates of return and TFP growth have declined in recent years compared to the long run trends.

Another contribution is to document what can happen to user costs when ex post asset inflation rates are used in the user cost formula. Dale Jorgenson and his coworkers have advocated the use of ex post inflation rates in a user cost formula and so we call the resulting user costs "Jorgensonian". We show that for many assets, Jorgensonian user costs can be quite volatile and even negative at times which means that they cannot be used in many contexts. We advocate the use of predicted asset inflation rates in the user cost formula and we suggest a very simple moving average method for forming these predicted asset inflation rates, which we implement and compare with their Jorgensonian counterparts. We use Jorgensonian and predicted user costs to construct alternative measures of capital services and TFP growth for our two sectors of the US economy and, somewhat surprisingly, we find that there was little difference in the resulting trend measures of TFP growth, even though there are very large differences in the two sets of user costs.

An additional contribution is the examination of what happens to ex post rates of return on assets employed and on TFP growth as we withdraw assets from the asset base. This research has relevance for existing estimates of rates of return and TFP growth since many productivity studies exclude land and inventories from their asset base. We find that excluding these assets leads to exaggerated estimated rates of return on the remaining assets (as could be expected) but the effects on estimates of TFP growth are more variable. For our Sector 1, we found that excluding land and inventories had little effect on measured TFP growth but in Sector 2, the exclusion of land dramatically lowered measured TFP growth.

Finally, we use our data set to provide evidence on the debate regarding growing inequality due to a falling labour share in income. We find that moving from value added shares to (Hayekian) income shares provides stronger evidence of falling labour shares, indicative of growing inequality, for both our sectors.

Our accounting framework is laid out in the following section and the empirical results for the above measurement exercises follow in the subsequent sections.¹

¹The Appendix in Diewert and Fox (2016) explains in detail how we used the Integrated Macroeconomic Accounts to construct our data set for the two sectors of the US economy.