Economic Modelling 27 (2010) 641–647

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Threshold effects in panel data stochastic frontier models of dairy production in Canada

Clément Yélou [a], Bruno Larue [a,*], Kien C. Tran [b]

[a] *Center for Research on the Economics of Agrifood (CREA), Université Laval, Canada*
[b] *Department of Economics, University of Lethbridge, Canada*

ARTICLE INFO

Article history:
Accepted 5 January 2010

JEL classification:
C12
C13
C23
C52

Keywords:
Stochastic frontier models
Threshold regression
Technical efficiency
Bootstrap
Dairy production

ABSTRACT

One of the most enduring problems in econometrics is how to properly account for heterogeneity among firms. Threshold regression models are intuitively appealing methods to deal with this issue. We consider a fixed-effect panel data stochastic frontier model (Schmidt and Sickles, 1984; Martin-Marcos and Suarez-Galvez, 2000) and, relying on Hansen (1999, 2000a), we propose an estimator that accommodates multiple thresholds. Our model assumes absence of any unmeasured time invariant heterogeneity across firms as in Greene (2005, p. 277). Slope and threshold parameters can be estimated using a within estimator combined with a grid search over the threshold parameters. Testing for threshold effects is problematic because threshold parameters are not identified under the null hypothesis, a case of the so-called Davies' problem. We apply the bootstrap procedure proposed by Hansen (1999, 2000a) to test for the presence of thresholds. An asymptotic confidence set for the threshold parameter can be obtained by inverting an LR test, using the distribution result presented in Hansen (1999, 2000a). Our empirical application features a panel of Quebec dairy farms. We use farm size as the threshold variable. The presence of a trend in the specification matters for the determination of the number of thresholds. Technical efficiency scores and rankings of farms estimated from competing model specifications are highly correlated and do not vary significantly across groups of farm sizes defined by the threshold parameter values.

1. Introduction

Structural change and threshold effects are two related issues that have motivated considerable empirical and theoretical research in time series econometrics (e.g. Tsay, 1989, 1998; Enders and Granger, 1998; Hansen, 2000a,b). An important problem in cross-sectional or panel data models is heterogeneity among firms. One approach to address this issue is to compare a regression function that is identical across all observations in a sample to a set of regression functions that allow for observations to fall into discrete classes as in Hansen (1999). With a focus on the dairy sector, we implement a stochastic frontier technical efficiency analysis that controls for heterogeneity among firms using a multiple-regime threshold production function.

Threshold regression models are intuitively appealing econometric methods to account for heterogeneity. In the context of stochastic production frontier models, the question may be whether larger firms use a production technology that differs from that of smaller firms. This would allow researchers to determine whether the higher productivity of large firms stems from the use of a different technology or simply a more efficient use of inputs given the constraints imposed by a common technology (see Tran and Tsionas, 2006). Related methods that allow for heterogeneity in stochastic frontier models include latent class models (Greene, 2002, 2005; Orea and Kumbhakar, 2004), random coefficients models (Tsionas, 2002; Greene, 2002, 2005) and Markov switching frontier models (Tsionas and Kumbhakar, 2004). The distinguishing feature of threshold models is that they assume that heterogeneity is induced by an observable exogenous variable, as opposed to unobservable random terms.

Recently, Tsionas and Tran (2006) have proposed various models that allow for heterogeneity in technology and in the distribution of technical inefficiency. Bayesian inference methods are proposed for the estimation of these models and for model comparisons. Bayesian tools such as the posterior odds ratio and the Bayes factor are proposed for model selection, including the comparison of a threshold model against a model without threshold effects. These statistics are used as evidence pertaining to the presence of threshold effects in the data. However, from a classical inference approach, such evidence needs to be based on a statistical test. Testing for threshold effects is problematic and requires non-standard tools because of the presence of a nuisance parameter which is not identified under the null hypothesis. This is known as Davies' problem and appropriate

* Corresponding author. Paul-Comtois Bldg, 2425 de l'Agriculture St., Room 4417, Quebec City, QC, Canada G1V 0A6. Tel.: +1 418 656 2131x5098; fax: +1 418 656 7821.
E-mail addresses: clement.yelou@gmail.com (C. Yélou), bruno.larue@eac.ulaval.ca (B. Larue), kien.tran@uleth.ca (K.C. Tran).

0264-9993/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2010.01.003

techniques have been proposed in Davies (1987), Andrews (1993) and Hansen (1996, 1999, 2000a). For our specific threshold effects problem, the nuisance parameter is the value of the threshold. In this paper, we consider a simple threshold stochastic frontier model. Relying on Hansen (1999, 2000a), we propose an estimation method using a grid search over the threshold parameter, and we provide a method to test the presence of one or more thresholds in a parametric stochastic frontier model with panel data. We consider a panel data stochastic frontier model with fixed effects time invariant inefficiency term. That is, we assume that the technical inefficiency term is a firm-specific constant, as in the fixed effects and random effects panel data models of Schmidt and Sickles (1984), Horrace and Schmidt (1996), Greene (1997) and Martin-Marcos and Suarez-Galvez (2000). However, because the statistical inference methods we use were specifically proposed for fixed effects panel data models, we do not consider the stochastic frontier models with random inefficiency terms. This formulation of the panel data stochastic frontier model has the advantage that it does not require any distributional assumption for technical inefficiency.

The paper illustrates that methods discussed in Hansen (1999, 2000a) have wide-ranging empirical applications in the analysis of technical efficiency using panel data. We report results from an empirical application that involves a panel of 302 dairy farms located in the province of Quebec and observed during 11 years, over the period 1993–2003. For this application, the threshold variable is the number of cows, a proxy for farm size.

The rest of the paper is organized as follows. Section 2 first describes how the milking technology used in dairy production in Quebec may give rise to heterogeneity among dairy farms, and then sets the statistical framework for the threshold stochastic frontier model. The estimation method is presented in Section 3 while Section 4 describes the method to test for the presence of thresholds. Section 5 focuses on the methodology proposed to construct a confidence set for the threshold parameters. Section 6 presents the results from our empirical application. The concluding section summarizes our results and their policy implications.

2. Thresholds in dairy production and statistical framework

This section first describes how threshold effects may arise in the dairy production technology and then presents our statistical framework for estimating technical efficiency scores.

2.1. Input lumpiness, input fixity and thresholds in dairy production

Lumpiness and fixity are common traits of inputs used in agriculture and dairy production in Canada is no exception. For example, dairy production in Canada requires heated dairy barns and the whole barn must be heated whether the barn is completely filled or half filled. Once the barn is filled and the maximum number of cows is reached, dairy production can only be increased through higher output per cow which can be achieved by increasing other inputs. Naturally, barns come in different sizes, but to the extent that there are "common" sizes and that small changes in capacity are relatively too costly to be worth doing, we can expect that most dairy farms operate close to their capacity limits and that increases in size from a constrained equilibrium come in large lumps. Similarly, other inputs, like tractors, can provide various levels of services up to a maximum capacity. Moreover, different milking systems are used in dairy production. Because of their small size, many dairy herds in Quebec and Ontario, the two largest dairy producing provinces in Canada, are housed in tie stalls. This technology is labor intensive, but it does not require a large capital investment. Larger farms tend to use different variations of free stalling. In most cases, milking is done at fixed hours twice every day. However, new voluntary milking systems are being used by larger dairy farmers.[1] Clearly, there are lumps and capacity constraints associated with milking systems.

Typical estimation of technical efficiency scores relies on a homogenous production function to define the efficiency frontier. The above discussion clearly shows that the lumpiness, capacity constraints and fixity of several key inputs introduce heterogeneity in milk production and this heterogeneity is intuitively driven by the size of the dairy herd.

2.2. Statistical framework

We consider the following two-regime threshold stochastic frontier production model for the dairy sector:

$$y_{it} = \alpha + \beta_1' x_{it} I(q_{it} \le \gamma) + \beta_2' x_{it} I(q_{it} > \gamma) - u_i + v_{it},\ u_i \ge 0, \tag{1}$$

where for firm i at time period t, $i=1, \ldots, N$, $t=1, \ldots, T$, y_{it} is the logarithm of output, $x_{it} \in \mathbb{R}^k$ is a vector of logarithm of inputs, $I(.)$ is the indicator function, β_1 and β_2 are two vectors of parameters associated with two different technologies Γ_1 and Γ_2, respectively; v_{it} is statistical error term, and $u_i \ge 0$ represents technical inefficiency. We assume throughout that the error term v_{it} is independent and identically distributed with mean zero and finite variance σ_v^2. q_{it} is an exogenous and observable threshold variable that governs the technology regime of firms in that sector; γ is the threshold value such that at time t, firms for which $q_{it} \le \gamma$ adopt technology Γ_1 whereas all the other ones adopt technology Γ_2. To motivate this formulation with threshold effects, consider some stylized facts about dairy production. Even though there is a high proportion of small dairy farms, not all of the farms use the same milking system. Some farms are large enough to mix their feed on the farm. Some have little land or are located in areas where it is difficult to produce corn. Hence, it is justified to entertain the possibility that farms need not have the exact same technology. We may therefore posit that technological jumps occur at various farm sizes and accordingly production frontier models with one or more thresholds may be considered. For $\beta_1 = \beta_2$, we get the basic panel data stochastic frontier model (see Pitt and Lee, 1981; Schmidt and Sickles, 1984; Cornwell and Schmidt, 1995; Greene, 1997; Martin-Marcos and Suarez-Galvez, 2000). An assumption that underlies our threshold effects formulation is that firms elect to switch from the Γ_1-technology to the Γ_2-technology because the cost of the new technology is lower than the loss in profit that firms would incur if they were not to switch. The time invariance assumption for the technical inefficiency term u_i may be an unreasonable one in long panels; we refer to Kumbhakar (1990) where it is argued that firms aware of their relative inefficiency would take steps to catch-up over time. This is less of a concern in our sample because all of the farms belong to management clubs which makes for more stable individual efficiency levels and rankings. Furthermore, the time invariant technical efficiency frontier model has a long history in the panel data stochastic frontier literature and it can be easily integrated in a multiple threshold framework. As Kumbhakar and Lovell (2000, p. 99) put it, the fixed effects model has the virtue of simplicity and nice consistency properties. For further reference, see Pitt and Lee (1981), Schmidt and Sickles (1984), Greene (1997), Horrace and Schmidt (1996), Martin-Marcos and Suarez-Galvez (2000), Kim et al. (2006) and references therein.

[1] These systems were first introduced in Europe, where grazing space is more limiting. Basically, the cows stay inside barns and get milked with robotic technology at the time they choose. The typical voluntary milking unit in Europe is for 60–70 cows (Wikipedia, 2008). In Ontario, the average user has 94 cows, even though such a system reportedly allows the family farm to expand up to 150 cows without hiring outside labor (Ontario Ministry of Agriculture, Food and Rural Affairs, 2008).