

RISE | RICE INITIATIVE *for the* STUDY *of* ECONOMICS

RISE Working Paper 15-008

“Semi-nonparametric Spline Modifications to the Cornwell-Schmidt-Sickles Estimator: An Analysis of U.S. Banking Productivity ”

by

Pavlos Almanidis, Giannis Karagiannis and Robin C. Sickles



RICE

Department of Economics

Baker Hall, MS22

6100 Main Street, Houston, Texas 77005

<https://economics.rice.edu>

Semi-nonparametric Spline Modifications to the Cornwell-Schmidt-Sickles
Estimator: An Analysis of U. S. Banking Productivity

Pavlos Almanidis¹
International Tax Services
Ernst&Young LLP
Toronto, Canada
pavlos.almanidis@ca.ey.com

Giannis Karagiannis
Department of Economics
University of Macedonia
Thessaloniki, Greece
karagian@uom.gr

Robin C. Sickles
Department of Economics
Rice University
Houston, Texas
rsickles@rice.edu

This version June, 2015

Keywords: Cornwell-Schmidt-Sickles Estimator; Time-varying efficiency; Spline
Functions; Semi-parametric estimation

JEL classification codes: C13, C21, C23, D24, G21.

¹ The views expressed by the first author are independent of those of Ernst & Young LLP.

Semi-nonparametric Spline Modifications to the Cornwell-Schmidt-Sickles Estimator: An Analysis of U. S. Banking Productivity

This paper modifies the Cornwell, Schmidt and Sickles (CSS) (1990) time-varying specification of technical efficiency to allow for switching patterns in temporal changes, which may occur more than once during the sample period. For this purpose, we move from the (second-order) polynomial specification of the standard CSS to a spline function set up while keeping CSS's flexibility regarding the cross section dimension. The spline function specification of the temporal pattern of technical efficiency can accommodate more than one turning point and thus can be non-monotonic. This allows the modeler to account for firm or individual efficiency gains that can occur relatively quickly, for example, changes related to regulation or policy changes as well as those related to ownership/organization changes (e.g., merger or acquisitions).

1. Introduction

One of the interesting aspects of performance evaluation analysis is dynamic benchmarking, namely identifying and estimating the temporal patterns of efficiency per se and its role as an inherent component of productivity growth. In many cases, the effect of efficiency changes has been found as important as technical change in determining the evolution of productivity growth. One such example is the case of Japan during the period 1979-1988 as illustrated in Fare *et al.* (1994) where changes in technical efficiency and the catching-up process that resulted was found to be the most important source of growth for aggregate labor productivity.

Some will argue, however, that the accuracy of such empirical findings may depend on how time-varying technical efficiency has been modeled. Usually a linear time trend is used to capture the time pattern of efficiency changes (e.g. Kumbhakar, 1990; Battese and Coelli, 1992; Cuesta, 2000). This is a rather restrictive formulation of time-varying efficiency as its changes over time are given by a constant rate. That is, efficiency is either increasing or decreasing at a constant rate. It is also common to assume that the time pattern of technical efficiency is uniform for all producing units in the sample or the firms in the industry. Even though the assumption of a common temporal pattern is restrictive is not unreasonable for putty-clay type industries.

On the other hand, several flexible specifications have been proposed in the literature for modeling time-varying efficiency, such as those of Cornwell, Schmidt and Sickles (1990), Kumbhakar (1990), Lee and Schmidt (1993), Lee (2006), Karagiannis and Tzouvelekas (2007) Ahn, Lee and Schmidt (2007, 2013), and Kneip, Sickles, and Song (2012). Among them the Cornwell, Schmidt and Sickles (1990) (CSS) is still one of the more flexible specifications of temporal variation in efficiency, it is relatively easy to implement, and can accommodate settings in which technical inefficiency and the inputs or other regressors (i.e., environmental variables) can be correlated.² Although the specification used in the CSS analysis allows for firm-specific patterns of time-varying efficiency that can change through time by means of a quadratic function of time, such a specification was chosen for purposes of illustration and due to erratic behavior of higher order terms.

The aim of this paper is to generalize the CSS time-varying specification to allow for more erratic patterns of temporal changes, which in turn will allow for multiple turning points. For this purpose, we specify a spline function set up while keeping CSS's flexibility in the cross-section dimension. The generalization puts more emphasis on firm heterogeneity in terms of growth rates rather than level differences in efficiency. The spline function specification of the temporal pattern of technical efficiency can also be non-monotonic due to periods in which firms may face radical regulation or policy changes as well as shocks related to changes in ownership/organization (e.g., merger or acquisitions).

In the next section we briefly discuss the general productivity model we employ and decompositions into technology and efficiency change that can be made with it. We then briefly outline the CSS estimator that has been used in many applications to measure such important aspects of economic growth and point out its generality in addressing environmental effects, an attribute of the estimator that has apparently been missed by many researchers. In Section 3 we present a spline modification of the CSS estimator and discuss a semi-parametric method for its estimation. Section 4 discusses the banking data and the empirical results from the illustration we use to introduce the new estimator, while section 5 concludes.

² Much more general mixed types of so called "environmental" variables can also be controlled with the CSS model such as additive effects that impact the slope coefficients of the "environmental" variables. In these specifications of the CSS estimator the "environmental" variables impact the frontier as well as the level of efficiency, unlike most two step models wherein there is separability between the frontier and efficiencies of the cross-section units, such as firms (Wang, 2002; Wang and Schmidt, 2002; Simar and Wilson, 2007).

2. Econometric Specification of the Productivity Model and the CSS Estimator

Regression based approaches to decompose productivity growth into technical change and efficiency change components can be based on a rather straightforward generic model of production using a multiple output / multiple input technology specified by means of the input or the output distance function (Caves, Christensen and Diewert, 1982; Coelli and Perelman, 1999; Sickles, Good, and Getachew, 2002). Distance functions that are linear in parameters, such as the linear in logs Cobb-Douglas or translog or linear in levels generalized-Leontief or quadratic, constitute the predominant functional forms used in productivity studies. Treatments for unobserved technical efficiency (heterogeneity) can be motivated with the following classical model for a single output technology estimated with panel data assuming unobserved firm effects:

$$y_{it} = x_{it}\beta + \eta_i(t) + v_{it}$$

where $\eta_i(t)$ represents the unit specific fixed effect that may be time varying, x_{it} is a vector of regressors, some of which may be endogenous and correlated with the error v_{it} or the effects $\eta_i(t)$. This is the basic regression model for single output technologies used in many empirical studies.

We start with a relatively simple representation of the output distance function as an m - output, n -input deterministic distance function $D_I(Y, X)$ given by the Young index, described in Balk (2008):

$$D_I(Y, X) = \frac{\prod_{k=1}^n X_{kit}^{\delta_k}}{\prod_{j=1}^m Y_{jit}^{\gamma_j}}$$

The input-distance function is non-increasing, linear homogeneous in inputs, convex in X and non-decreasing and quasi-convex in Y . After taking logs, adding a disturbance term v_{it} to account for nonsystematic error in observations, functional form mis-specifications, etc. and a technical efficiency term $\eta_i(t)$ to reflect the nonnegative difference between the upper bound of unity for the distance function and the observed value of the distance function for firm i at time t , we can write the input distance function as:

$$-x_{1,it} = \sum_{j=1}^m \gamma_j y_{jit}^* + \sum_{k=2}^n \delta_k x_{kit}^* + \eta_i(t) + v_{it}$$

where $y_{jit}^* = \ln(Y_{jit} / Y_{1it})$ and $x_{kit}^* = \ln(X_{kit} / X_{1it})$. After redefining a few variables the distance function can be written in the canonical form as

$$x_{it}^* = z_{it} \beta + \eta_i(t) + v_{it}.$$

where $z_{it} = [x_{it}^* \ y_{it}^*]$. The functional form utilized in most studies is the Cobb-Douglas specification. Even though it has been criticized for its assumption of separability of outputs and inputs, as pointed out by Coelli (2000) among many others, the Cobb-Douglas remains a reasonable and parsimonious first-order local approximation to the true function. The translog input distance function, where the second-order terms allow for greater flexibility, proper local curvature, and lift the assumed separability of outputs and inputs, can also be framed in this canonical model representation of a linear panel model with unit-specific and time-varying heterogeneity. The translog input distance function takes the form:

$$\begin{aligned} -y_{lit} = & \sum_{j=2}^m \gamma_j y_{jit}^* + \frac{1}{2} \sum_{j=2}^m \sum_{l=1}^m \gamma_{jl} y_{jit}^* y_{lit}^* + \sum_{k=1}^n \delta_k x_{kit}^* + \frac{1}{2} \sum_{k=1}^n \sum_{p=1}^n \delta_{kp} x_{kit}^* x_{pit}^* \\ & + \sum_{j=2}^m \sum_{k=1}^n \theta_{jk} y_{jit}^* x_{kit}^* + \eta_i + u_{it} \end{aligned}$$

Since the model is linear in parameters, then after redefining a few variables the translog distance function also can be written as³

$$x_{it}^* = z_{it} \beta + \eta_i(t) + v_{it}.$$

Of course, if the technology involves multiple outputs, then the right hand side endogenous variables must be instrumented. Whether or not the effects need to be instrumented depends on their orthogonality with all or a subset of the regressors.

This is the model vehicle we use for estimating efficiency change using the frontier methods we specify below. If we assume that innovations are available to all firms and that firm-specific idiosyncratic errors are due to relative inefficiencies then we can decompose sources of *TFP* growth in a variety of ways. The overall level of innovation change (innovation is assumed to be equally appropriable by all firms) can

³ A similar transparent reparametrization of any distance function that is linear in parameters can be used to estimate other linear in parameters distance functions such as the generalized Leontief or the quadratic.

be measured directly by such factors as a distributed lag of R&D expenditures, or patent activity, or some other direct measure of innovation. It can be proxied by the time index approach of Baltagi and Griffin (1988), linear time trends, or some other type of time variable. Innovation measured in any of these ways would be identified under non-pathological circumstances. Direct measures can be identified under an assumption that the matrix of regressors has full column rank, and the indirect measures can be identified by functional form assumptions. For example, the index number approach used in Baltagi and Griffin is identified by its nonlinear construction. Innovation can also be proxied by exogenous or stochastic linear time trends (Bai, Kao and Ng, 2009), which are often identified by nonlinear specifications of time varying inefficiency used in the models of Battese and Coelli (1992) and Lee and Schmidt (1993). Other types of restrictions can be employed, such as the orthogonality conditions utilized in the Cornwell, Schmidt, and Sickles (1990) estimator we extend in our analysis using cubic splines.

The Cornwell Schmidt and Sickles (CSS) (1990) panel stochastic frontier model itself extends the basic panel data model of Pitt and Lee (1981) and Schmidt and Sickles (1984) to allow for heterogeneity in slopes as well as intercepts. Thus, in the model $y_{it} = x_{it}\beta + \eta_i(t) + v_{it}$ the effects are specified as $\eta_i(t) = W_{it}\delta_i + v_{it}$. The L coefficients of W , the terms in the vector δ_i , depend on different units representing heterogeneity in slopes. In their application to the US commercial airline industry, CSS specified $W_{it} = (1, t, t^2)$, although this was intended by the authors' to be a parsimonious parameterization useful for their application. It does not in general limit the effects to be quadratic in time.

A common construction can relate this model to standard panel data model. Let $\delta_i = \delta_0 + u_i$ and $\delta_0 = E[\delta_i]$. Then the model can be written as:

$$\begin{aligned} x_{it}^* &= Z_{it}\beta + W_{it}\delta_0 + \varepsilon_{it}, \\ \varepsilon_{it} &= W_{it}'u_i + v_{it}. \end{aligned}$$

Here u_i are assumed to be i.i.d. zero mean random variables with covariance matrix Δ . The disturbances v_{it} are taken to be i.i.d. random variable with a zero mean and constant variance σ^2 , and uncorrelated with the regressors and u_i . In matrix form the model is:

$$\begin{aligned}x^* &= Z\beta + W\delta_0 + \varepsilon, \\ \varepsilon &= Qu + v\end{aligned}$$

where $Q = \text{diag}(W_i)$, $i = 1, \dots, N$ is a $NT \times NL$ matrix, and u is the associated $NL \times 1$ error vector.

Three different estimators can be derived based on differing assumptions regarding the correlation of the efficiency effects and the regressors, specifically, the correlation between the error term u , and regressors Z and W . These are the *within* and GLS, which we employ in this paper, and *the efficient IV* estimator.⁴ We briefly discuss below the *within* and GLS estimators, which we will modify using the spline extensions in the following section.

The *within* estimator allows for correlation between all of the regressors and the effects. Let $P_Q = Q(Q^{-1}Q)$, $M_Q = I - P_Q$. Then the CSS *within* (CSSW) estimator of β is given by:

$$\hat{\beta}_w = (Z'M_Q Z)^{-1} Z'M_Q x^*.$$

The GLS estimator is consistent when no correlation exists between the technical efficiency term and the regressors, as in Pitt and Lee (1981), Schmidt and Sickles (1984) and many others that utilize this standard random effects assumption. The variance of the composed error is given by

$$\text{cov}(\varepsilon) = \Omega = \sigma^2 I_{NT} + Q'(I_N \otimes \Delta)Q.$$

CSS show that

$$\Omega^{-1/2} = \frac{1}{\sigma} M_Q + F$$

where $F = Q(Q^{-1/2}[\sigma^2 I_{NL} + (Q^{1/2}(I_N \otimes \Delta)(Q^{1/2})^{-1/2}(Q^{-1/2}Q')])^{-1/2}(Q^{-1/2}Q')$. The transformed model is thus $\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}W\delta_0 + \Omega^{-1/2}\varepsilon$. CSS provide formulae for the feasible consistent estimates of $\Omega^{-1/2}$.

For either the *within* or GLS estimators the δ_0 are estimated by regressing the residuals for firm i on $W_{it} = (1, t, t^2)$ and the fitted values from this regression provide consistent ($T \rightarrow \infty$) estimates of the $E[\eta_i(t)]$. This is analogous to the approach in Schmidt and Sickles (1984) when there is no temporal variation in the unit specific technical efficiencies. Relative efficiencies, normalized by the consistent estimate of

⁴ Details on the efficient IV estimator can be found in the CSS paper.

the order statistics identifying the most efficient unit, are then calculated as $\widehat{\eta}(t) = \max_j[\widehat{\eta}_j(t)]$. Efficiency of firm i relative to the most efficient firm is then given by $RE_i(t) = \widehat{\eta}(t) - \widehat{\eta}_i(t)$.⁵

3. Spline Model Specification

Another flexibility aspect in the specification of technical efficiency involves monotonicity over time. In many previous specifications (e.g., Battese and Coelli, 1992; Cuesta, 2000), the effect of the passage of time on technical inefficiency is necessarily monotonic and thus may be either efficiency-enhancing or efficiency-impeding, but not both. Others, such as Lee and Schmidt (1993) and Lee (2006) allow for more general patterns. We consider below the quadratic spline as a special case of restricted least squares and thus once the number of knots is set (or tested for sequentially) a general spline estimator can be specified (Buse and Lim, 1977). Depending on the number of break (turning) points, which may be determined by either prior information regarding the sector under consideration (Bottasso and Conti, 2009) or by the process suggested by Fox (1998), the time pattern can be rather flexible, curved or monotonic. The latter two options can be tested statistically as nested model specifications.

The Diewert-Wales (1992) quadratic spline function can be incorporated into the CSS specification in order to obtain a flexible and parsimonious specification of the temporal pattern of technical efficiency, allowing more than one turning point. This specification allows for firm-specific patterns of temporal variation of technical efficiency and captures effects not visible in those models that assume a common pattern of technical efficiency. In addition, we can test (i) for the existence of a common temporal pattern for all firms in the sample as well as (ii) the hypothesis of time-varying technical efficiency for all or some of the firms in the sample.

⁵ Firm-specific relative efficiencies can be identified along with the overall growth in innovation that diffuses to all firms for the GLS estimator. Under appropriate orthogonality assumptions, a similar term can be identified for the Hausman-Taylor type efficient IV estimator. Thus for these two estimators total factor productivity can be decomposed into technical change and efficiency change. Such a decomposition is not possible for the CSS *within* estimator as the technical change term is not identified after the within transformation.

The spline extension of the CSS model can be introduced by assuming a single time break $t_1 \in [0, T]$ in the inefficiency level of firm i . An extension to a model with multiple time breaks is straightforward and is discussed below. Following Diewert and Wales (1992), the inefficiency function of the CSS model can be represented by a quadratic spline function as follows:

$$u_{it} = \begin{cases} \delta_{i0} + \delta_{i1}t + \delta_{i2}^{(1)}t^2 & \text{if } t \leq t_1 \\ \delta_{i0} + \delta_{i1}t_1 + \delta_{i2}^{(1)}t_1^2 + (t - t_1)(\delta_{i1} + 2\delta_{i2}^{(1)}t_1) + (t - t_1)^2\delta_{i2}^{(2)} & \text{if } t > t_1 \end{cases}$$

Here the superscripts on the quadric term parameters relate to the periods before and after the time break point t_1 .

The above can be rewritten as

$$u_{it} = \begin{cases} \delta_{i0} + \delta_{i1}t + \delta_{i2}^{(1)}t^2 & \text{if } t \leq t_1 \\ \tilde{\delta}_{i0} + \tilde{\delta}_{i1}t + \tilde{\delta}_{i2}t^2 & \text{if } t > t_1 \end{cases}$$

where

$$\begin{aligned} \tilde{\delta}_{i0} &= \delta_{i0} + (\delta_{i2}^{(2)} - \delta_{i2}^{(1)})t_1^2 \\ \tilde{\delta}_{i1} &= [\delta_{i1} + 2(\delta_{i2}^{(1)} - \delta_{i2}^{(2)})t_1] \\ \tilde{\delta}_{i2} &= \delta_{i2}^{(2)} \end{aligned}$$

Clearly, the above model specification allows for the levels and slopes of inefficiencies to differ across these two time periods. Notice that, if $\delta_{i2}^{(1)} = \delta_{i2}^{(2)} = \delta_{i2}$ then the above model collapses to the standard CSS model with no breaks. Also, it is worthwhile to note that both the inefficiency function and its first derivative with respect to time are continuous at t_1 . The continuity feature is crucial here as it allows for a smooth transition from one state to another (no jumps).

In order to proceed with the estimation of the CSS model with the quadratic spline specification, let first the matrix of time regressors to be denoted as follows:

$$W_i = \begin{bmatrix} W_{1i} & 0 \\ 0 & W_{2i} \end{bmatrix}$$

where $W_{1i} = [1, t, t^2]I(t \leq t_1)$, $W_{2i} = [1, t, t^2]I(t > t_1)$ and $I(\cdot)$ is an indicator function. Subsequently, define $Q = \text{diag}(W_i)$ and $M_Q = I - Q(Q'Q)^{-1}Q'$ as a projection

matrix onto column space of Q . Then while fixing t_1 , the within estimator of the structural parameters is given by:

$$\hat{\beta}_w = (X'M_Q X)^{-1} X'M_Q y$$

and the concentrated sum of squared errors, which is a function of the observed data and time, is given by:

$$S(t) = \hat{e}'\hat{e} = (y - X(X'M_Q X)^{-1} X'M_Q y)'(y - X(X'M_Q X)^{-1} X'M_Q y)$$

where $\hat{e} = y - X\hat{\beta}_w$ represents the within residuals. The time break t_1 is estimated by minimizing $S(t)$ using grid search techniques. That is,

$$\hat{t}_1 = \arg \min_t S(t)$$

Once \hat{t}_1 is estimated, which represents the least square estimator of t_1 , the slope coefficient is obtained and the residual variance is estimated by dividing $S(\hat{t})$ by the degrees of freedom, $N(T-1)$.

The model with a single time break can be extended to accommodate multiple time breaks. A model with k time breaks can be represented as follows:

$$u_{it} = \begin{cases} \delta_{i0} + \delta_{i1}t + \delta_{i2}^{(1)}t^2 & \text{if } t \leq t_1 \\ \delta_{i0} + \delta_{i1}t_1 + \delta_{i2}^{(1)}t_1^2 + (t_2 - t_1)(\delta_{i1} + 2\delta_{i2}^{(1)}t_1) + (t_2 - t_1)^2 \delta_{i2}^{(2)} & \text{if } t_1 < t \leq t_2 \\ \dots & \dots \\ \delta_{i0} + \delta_{i1}t_{k-1} + \delta_{i2}^{(k-1)}t_{k-1}^2 + (t_k - t_{k-1})(\delta_{i1} + 2\delta_{i2}^{(k-1)}t_k) + (t_k - t_{k-1})^2 \delta_{i2}^{(k)} & \text{if } t_{k-1} < t \leq t_k \\ \delta_{i0} + \delta_{i1}t_k + \delta_{i2}^{(k)}t_k^2 + (t - t_k)(\delta_{i1} + 2\delta_{i2}^{(k)}t_k) + (t - t_k)^2 \delta_{i2}^{(k+1)} & \text{if } t > t_k \end{cases}$$

Similar to the single time break model, the multiple time breaks are estimated via a grid search algorithm. Joint estimation of the time breaks requires a grid search over a large number of time break combinations. Therefore, the time breaks are sequentially estimated as suggested in Hansen (1999). The sequential estimation of the threshold parameters is consistent and is more than necessary especially when the time period under consideration is long. A drawback of the sequential estimation method is that it yields asymptotically efficient estimates only for the last time break in the estimation process. The previous estimates are contaminated by the presence of the neglected time breaks. We follow Bai (1997) and utilize a refinement estimation of the time breaks parameters, which amounts to re-estimating the time break parameters backwards, each time holding the estimates of the previous time breaks fixed.

It is important to test whether the time breaks are statistically significant or

not. However, the distribution of t_1 is nonstandard which complicates inference. In particular, testing for the presence of a time break becomes problematic, since t_1 is not identified under the null hypothesis of no threshold and conventional tests would have distributions that are also nonstandard (the Davies' Problem, 1977). Following Hansen (1999), we utilize a bootstrap method to simulate the asymptotic distribution of the classical LR test in our inference on time break estimates.

The bootstrap testing is carried out in the following steps:

1. Estimate the model under the null and alternative hypotheses and calculate the LR statistic as $LR = (S_0 - S_1) / \sigma^2$
2. Estimate the sample of residuals $\hat{\epsilon}$ under the null hypothesis and treat this sample as the empirical distribution in bootstrap replications
3. Fix the data and draw (with replacement) a sample of size N from the empirical distribution above and use these errors to create a bootstrap sample
4. Using the bootstrap sample, estimate the model under the null and alternative hypotheses and calculate the bootstrap value of the likelihood ratio statistic LR^b
5. Repeat this procedure a large number of times and calculate the percentage of draws for which LR^b exceeds LR
6. Reject the null hypothesis if the percentage above exceeds the desired confidence level.

4. Data and Empirical Results

In this section we provide empirical evidence on a comparison between the two specifications of the CSS model based on the second-order polynomial and based on the spline function using a rather homogenous and balanced sample of large (too-big-to-fail) US banks. The data are from the quarterly consolidated reports on condition and income (Call Reports) for US commercial banks collected by the Federal Reserve Bank of Chicago and the Federal Deposit Insurance Corporation. The particular sample used in this study covers the period from the first quarter of 1984 to the third quarter of 2009 (i.e., 106 quarters) and refers to 45 banks with total asset size of at least US\$ 10 billion as of the second quarter of the 2010 fiscal year, with a total of 4,770 observations used in the estimation.

The analysis is based on the Sealey and Lindley (1977) intermediation approach according to which banks are viewed as financial intermediates that collect deposits and other funds to transform them into loanable funds by using capital and labor. In this case, deposits are viewed as inputs instead of outputs as in the production approach. We consider five outputs, namely real estate loans, commercial and industrial loans, loans to individuals, securities and off-balance sheet items. On the input side we have capital, labor, and interest-bearing deposits in total non-transaction accounts and purchased funds. Descriptive statistics for all model variables are given in Table 1.

We estimate the proposed spline specification of the CSS model using both the within and the GLS estimator as with the latter we can separate the effect of technical change from that of changes in technical efficiency even though both are modeled by time trend; this is an important aspect in the productivity decomposition analysis. We have also tried the Hausman-Taylor estimator in the case where some of the explanatory variables are not orthogonal with the “effects” capturing unobserved heterogeneity. Based on the Hausman-Wu test we have no evidence of simultaneity bias and thus the GLS estimator fits better with the data at hand. In addition and for comparison purposes we also estimate the conventional version (i.e., second-order polynomial) of the CSS model. The results are presented in Tables 1 and 2.

Estimation the proposed spline specification of the CSS involves both the determination of the unknown time breaks as well as the values of the structural model parameters. For this purpose, the proposed model is estimated by minimizing the concentrated sum of squared residuals using a grid search over possible time periods to determine in the first place the time breaks, as in Almanidis (2013), and then estimate the values of the structural parameters. In the estimation it is assumed that the timing of the breaks is the same for all banks but this does not necessarily mean that each bank will experience the break. This is apparent from the fact that efficiency is firm-specific in the CSS model. The estimated parameters of the model are presented in Table 2. Returns to scale are estimated on average to be decreasing while the positive sign of the time trend in the GLS model implies that technical change was progressive.

The estimated average efficiency is smaller with the spline specification than with the conventional specification of the CSS model. In particular, the average technical efficiency from the spline specification is around 64% for both within and

the GLS model while the corresponding figure for the conventional specification is at 70%. Besides that, there is no important differences in the cross sectional distribution of efficiency scores. There are however significant differences in the temporal pattern of efficiency scores.

Figure 1 depicts these differences in terms of average efficiencies. The estimated model revealed twelve time-break points based on the bootstrap test with 10,000 replications at each stage of time-break search. These time-break points occur in the third quarter of 1986, the second quarter of 1988, the third quarter of 1990, the fourth quarter of 1991 and 1994, the first quarter of 1997, the third quarter of 1999, and 2001, the second quarter of 2004, the third quarter of 2006, and the second quarter of 2007 and 2008. These points in the time correspond to the pre- and post-deregulation period of the U.S. banking industry, periods of technological and financial breakthroughs, as well as to periods of financial crises that affected the U.S. banking efficiency levels over the past three decades. The notable of these time points are: (i) the Federal Reserve's granting commercial bank holding companies with the power to underwrite corporate securities in 1987 (also allowing operating commercial banks to underwrite corporate securities in 1989); (ii) the savings and loan crisis of the early 1990s; (iii) the Reigle-Neal Interstate Banking and Branching Efficiency Act of 1994, which allowed the interstate banking and branching; (iv) the Gramm-Leach-Bliley Act Financial Services Modernization Act of 1999, which granted broad-based securities, investment, and insurance power to commercial banks; (v) the introduction of the internet banking and check clearing through imaging technology in early 2000s; (vi) the collapse of the U.S. housing market bubble in mid-2006 and the subsequent dramatic increase in delinquencies and default rates on subprime residential-mortgage-backed securities (RMBS); and (vii) 2007-2010 financial crisis.

Comparing with Figure 2, where the average efficiencies of the conventional CSS model are presented, it is evident the non-monotonic time pattern of the spline specification around the aforementioned turning points. In contrast, the average efficiencies predicted by the conventional specification reached a maximum around the end of 1994 and decreased thereafter.

5. Concluding Remarks

In this paper we present a specification of the CSS (1990) model that allows for more erratic patterns of temporal changes in technical efficiency. The model is based on a second order spline function which can accommodate more than one turning point over time. This non-monotonic temporal pattern depicts in a much more flexible way firm heterogeneity in terms of growth rates and it is particularly suitable for analyzing efficiency changes during periods of regulation or policy changes. We estimate such a model using a translog output distance function for a sample of large (too-big-to-fail) US banks using a semi-parametric approach and a grid search algorithm. We also used a bootstrap method to formally test for the presence of time breaks. Our empirical results reveal the presence of twelve time break points during the period 1984-2009 indicating a highly non-monotonic time pattern of average technical efficiency.

References

- Ahn, S.C., Lee, Y.H. and Schmidt, P. (2007) Stochastic Frontier Models with Multiple Time-varying Individual Effects, *Journal of Productivity Analysis*, 27, 1-12.
- Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013) "Panel Data Models with Multiple Time-varying Individual Effects." *Journal of Econometrics* 174(1):1-14.
- Almanidis, P. (2013) Accounting for Heterogeneous Technologies in the Banking Industry: A Time-varying Stochastic Frontier Model with Threshold Effects, *Journal of Productivity Analysis*, 39, 191-205.
- Balk, B.M. (2008) *Price and quantity index numbers: models for measuring aggregate change and difference*. New York: Cambridge University Press.
- Bai, J. (1997) Estimating multiple breaks one at a time. *Econometric Theory* 13, 315-352.
- Bai, J., C. Kao, and S. Ng. (2009) Panel cointegration with global stochastic trends. *Journal of Econometrics*, (149):82-99.
- Baltagi, B. and Griffin, J. (1988) A General Index of Technical Change, *Journal of Political Economy*, 96, 20-41.
- Battese, G.E. and T.J. Coelli (1992) Frontier Production Functions, Technical Efficiency and Panel Data: with Application to Paddy Farmers in India, *Journal of Productivity Analysis*, 3, 153-69.
- Bottasso, A. and M. Conti (2009) Piece Cap Regulation and the Ratchet Effect: A Generalized Index Approach, *Journal of Productivity Analysis*, 32, 191-201.
- Buse, A. and L. Lim (1977): Cubic Splines as a Special Case of Restricted Least Squares, *Journal of the American Statistical Association*, 72:357, 64-68.
- Caves, D.W., L.R. Christensen, W.E. Diewert (1982) The economic theory of index numbers and the measurement of input, output, and productivity, *Econometrica*, 1393-1414.
- Coelli, T. (2000), 'Econometric estimation of the distance function representation of a production technology', CEPA, University of New England, mimeo.
- Coelli, T., and S. Perelman (1999) A comparison of parametric and non-parametric distance functions: With Application to European Railways, *European Journal of Operational Research*, 117, 326-339.
- Cornwell, C., Schmidt, P. and R.C. Sickles (1990) Production Frontiers with Cross-sectional and Time-series Variation in Efficiency Levels, *Journal of Econometrics* 46: 185-200.
- Cuesta, R. (2000) A production model with firm-specific temporal variation in

technical inefficiency: with application to Spanish dairy farms. *Journal of Productivity Analysis*, 13(2):139-158.

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, 247-254.

Diewert, W.E. and T.J. Wales (1992) Quadratic Spline Models for Producers' Supply and Demand Functions, *International Economic Review*, 33, 705-722.

Fare, R., Grosskopf, S., Norris, M. and Z. Zhang.(1994) Productivity Growth, Technical Progress, and Efficiency Changes in Industrialized Countries, *American Economic Review*, 84, 66-83.

Fox, K.J. (1998) Non Parametric Estimation of Technical Progress, *Journal of Productivity Analysis* 10, 235-250.

Hansen, B.E. (1999) Threshold effects in non-dynamic panels: Estimation, testing, and inference, *Journal of Econometrics* 93, 345-368.

Karagiannis G. and V. Tzouvelekas (2007) A Flexible Time-varying Specification of the Technical Inefficiency Effects Model. *Empirical Economics* 33: 531-540.

Kneip, A., Sickles, R. C. and W. H. Song (2012) A new panel data treatment for heterogeneity in time trends, *Econometric Theory* 28, 590-628.

Kumbhakar, S.C. Production Frontiers, Panel Data and Time-varying Technical Inefficiency, *Journal of Econometrics*, 1990, 46, 201-12.

Lee, Y.H. A Stochastic Production Frontier Model with Group-specific Temporal Variation in Technical Efficiency, *European Journal of Operational Research*, 2006, 174, 1616-30.

Lee, Y. H., and P. Schmidt (1993). A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*, pages 237-255.

Pitt, M.M., L.F. Lee (1981) The measurement and sources of technical inefficiency in the Indonesian weaving industry, *Journal of Development Economics* 9, 43-64.

Schmidt, P. and R. C. Sickles (1984) Production Frontiers and Panel Data, *Journal of Business and Economic Statistics* 2, 1984, 367-374.

Sealey, J. and J.T. Lindly. Inputs, outputs, and a theory of production and cost of depository financial institutions, *Journal of Finance*, 1977, 32, 1251-66.

Sickles, R. C. (2013) "Productivity and Panel Data," with Hao, J., and Shang, C. Chapter 17 of *Oxford Handbook of Panel Data*, Badi Baltagi (ed.) New York: Oxford University Press, 2013.

Sickles, R. C., G. H. Good and L. Getachew (2002) Specification of distance functions using semi-and nonparametric methods with an application to the dynamic performance of eastern and western European air carriers," *Journal of Productivity Analysis, Special Issue in Honor of W. W. Cooper*, 17, S. Ray, P. Bogetoft, S. Kumbhakar, and J. Pastor (eds.), 133-155.

Simar, L. and P. Wilson (2007) Estimation and inference in two-stage, semi-parametric models of production processes, *Journal of Econometrics*, 136, 31-64.

Wang, H.J. (2002) Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model, *Journal of Productivity Analysis* 18: 241-253.

Wang, H.J. and P. Schmidt. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels, *Journal of Productivity Analysis*, 2002, 18, 129-44.

Table 1: Summary Statistics (mean and standard deviation) of Model Variables

	1984:Q1	1993:Q1	2000:Q1	2009:Q3
real estate loans	867,896 (2,027,284)	3,300,626 (6,149,823)	11,900,000 (26,400,000)	44,700,000 (89,100,000)
commercial & industrial loans	2,431,577 (6,560,456)	3,260,927 (6,707,246)	10,500,000 (22,100,000)	18,400,000 (36,600,000)
loans to individuals	593,755 (1,145,667)	1,625,214 (3,330,108)	3,902,677 (7,491,437)	11,600,000 (23,400,000)
securities	1,868,698 (3,627,331)	4,149,224 (5,668,735)	3,485,955 (7,679,057)	44,000,000 (96,200,000)
off-balance sheet items	2.31005 (5.75334)	1.62116 (1.68702)	1.36456 (1.3120)	1.03391 (1.05039)
total demand deposits	987,153 (1,509,827)	2025,806 (2,885,831)	4,332,390 (8,887,760)	8,138,886 (17,300,000)
total time & savings deposits	4,991,220 (11,900,000)	9,629,871 (18,800,000)	30,300,000 (59,800,000)	104,000,000 (214,000,000)
labor	4,164 (7,620)	6,014 (9,644)	13,563 (25,614)	23,781 (45,427)
Capital	104,298 (232,928)	253,391 (476,984)	611,758 (1,150,663)	1,231,292 (2,170,180)
purchased funds	2,027,599 (3,920,692)	3,943,412 (6,500,932)	13,000,000 (23,600,000)	13,000,000 (24,000,000)

Table 2: Estimated Model Parameters: Spline Specification

	Estimated Parameter	Standard Error	Estimated Parameter	Standard Error
	Within		GLS	
y ₁	-0.4742	0.0731	-0.5147	0.0131
y ₂	0.7517	0.0625	0.7313	0.0105
y ₃	-0.0325	0.0180	-0.0255	0.0051
y ₄	-0.0678	0.0377	-0.0368	0.0086
x ₁	-0.2232	0.0624	-0.2700	0.0158
x ₂	-0.9818	0.1479	-0.9654	0.0172
x ₃	-0.2173	0.1256	-0.2203	0.0221
x ₄	0.8685	0.1190	0.9024	0.0223
x ₅	0.2718	0.0379	0.2554	0.0108
y ₁ y ₁	0.0321	0.0050	0.0281	0.0010
y ₂ y ₁	0.0145	0.0031	0.0162	0.0007
y ₂ y ₂	-0.0700	0.0042	-0.0688	0.0007
y ₃ y ₁	0.0003	0.0008	0.0011	0.0002
y ₃ y ₂	0.0033	0.0008	0.0033	0.0002
y ₃ y ₃	-0.0072	0.0004	-0.0073	0.0001
y ₄ y ₁	-0.0027	0.0018	-0.0021	0.0005
y ₄ y ₂	-0.0030	0.0018	-0.0040	0.0004
y ₄ y ₃	0.0015	0.0006	0.0014	0.0002
y ₄ y ₄	0.0050	0.0015	0.0047	0.0005
x ₁ y ₁	-0.0125	0.0033	-0.0134	0.0011
x ₁ y ₂	-0.0019	0.0030	-0.0017	0.0008
x ₁ y ₃	0.0006	0.0011	0.0014	0.0003
x ₁ y ₄	-0.0058	0.0022	-0.0064	0.0007
x ₁ x ₁	0.0181	0.0043	0.0186	0.0016
x ₂ y ₁	0.0213	0.0086	0.0109	0.0018
x ₂ y ₂	0.0426	0.0076	0.0399	0.0014
x ₂ y ₃	-0.0098	0.0023	-0.0094	0.0006
x ₂ y ₄	-0.0149	0.0057	-0.0090	0.0012
x ₂ x ₁	-0.0504	0.0083	-0.0574	0.0025
x ₂ x ₂	-0.0721	0.0230	-0.0712	0.0047
x ₃ y ₁	0.0640	0.0107	0.0626	0.0023
x ₃ y ₂	0.0051	0.0089	0.0066	0.0016
x ₃ y ₃	-0.0037	0.0029	-0.0024	0.0008
x ₃ y ₄	0.0035	0.0058	-0.0014	0.0015
x ₃ x ₁	0.0334	0.0090	0.0363	0.0025
x ₃ x ₂	0.0009	0.0211	0.0023	0.0040
x ₃ x ₃	0.0694	0.0326	0.0765	0.0064
x ₄ y ₁	-0.1043	0.0091	-0.0952	0.0018
x ₄ y ₂	-0.0290	0.0065	-0.0279	0.0011
x ₄ y ₃	0.0048	0.0022	0.0032	0.0006
x ₄ y ₄	0.0132	0.0044	0.0120	0.0012
x ₄ x ₁	-0.0120	0.0071	-0.0141	0.0018
x ₄ x ₂	0.1344	0.0160	0.1409	0.0032
x ₄ x ₃	-0.0489	0.0220	-0.0580	0.0039

x ₄ x ₄	-0.0431	0.0186	-0.0398	0.0034
x ₅ y ₁	-0.0271	0.0039	-0.0244	0.0011
x ₅ y ₂	0.0151	0.0035	0.0144	0.0008
x ₅ y ₃	0.0040	0.0010	0.0036	0.0003
x ₅ y ₄	0.0021	0.0028	0.0026	0.0008
x ₅ x ₁	0.0163	0.0037	0.0197	0.0013
x ₅ x ₂	-0.0254	0.0080	-0.0256	0.0021
x ₅ x ₃	-0.0387	0.0074	-0.0406	0.0019
x ₅ x ₄	0.0045	0.0077	0.0056	0.0018
x ₅ x ₅	0.0554	0.0048	0.0519	0.0015
y ₁ t	-0.0029	0.0003	-0.0030	0.0000
y ₂ t	0.0015	0.0003	0.0014	0.0000
y ₃ t	0.0002	0.0001	0.0002	0.0000
y ₄ t	-0.0001	0.0001	0.0000	0.0000
x ₁ t	-0.0002	0.0002	-0.0003	0.0001
x ₂ t	-0.0003	0.0005	-0.0005	0.0001
x ₃ t	-0.0022	0.0006	-0.0024	0.0001
x ₄ t	0.0031	0.0004	0.0034	0.0001
x ₄ t	0.0006	0.0002	0.0007	0.0000
t	-	-	0.0001	0.0004

Note: $y_1 = \log(\text{commercial \& industrial loans/real estate loans})$, $y_2 = \log(\text{loans to individuals/real estate loans})$, $y_3 = \log(\text{securities/real estate loans})$, $y_4 = \log(\text{off-balance sheet items/real estate loans})$, $x_1 = \log(\text{total demand deposits})$, $x_2 = \log(\text{total time \& savings deposits})$, $x_3 = \log(\text{labor})$, $x_4 = \log(\text{capital})$, $x_5 = \log(\text{purchased funds})$ and $t = \text{time trend}$.

Table 3: Estimated Model Parameters: Polynomial Specification

	Estimated Parameter	Standard Error	Estimated Parameter	Standard Error
	Within		GLS	
y ₁	-0.3755	0.0971	-0.4443	0.0569
y ₂	0.4031	0.0896	0.4336	0.0459
y ₃	0.0459	0.0271	0.0465	0.0221
y ₄	0.2944	0.0649	0.3145	0.0374
x ₁	-0.4347	0.1085	-0.3536	0.0689
x ₂	-2.2830	0.2253	-1.7420	0.0746
x ₃	0.5044	0.1829	0.2103	0.0963
x ₄	-0.0252	0.1623	0.0415	0.0970
x ₅	0.2740	0.0667	0.2061	0.0472
y ₁ y ₁	-0.0183	0.0059	-0.0207	0.0045
y ₂ y ₁	0.0478	0.0043	0.0473	0.0029
y ₂ y ₂	-0.0998	0.0050	-0.0979	0.0031
y ₃ y ₁	-0.0122	0.0012	-0.0121	0.0010
y ₃ y ₂	0.0012	0.0010	0.0015	0.0008
y ₃ y ₃	-0.0094	0.0006	-0.0097	0.0004
y ₄ y ₁	-0.0097	0.0028	-0.0087	0.0023
y ₄ y ₂	-0.0156	0.0030	-0.0154	0.0019
y ₄ y ₃	-0.0003	0.0010	-0.0004	0.0009
y ₄ y ₄	0.0133	0.0029	0.0133	0.0022
x ₁ y ₁	0.0001	0.0057	0.0000	0.0046
x ₁ y ₂	-0.0008	0.0047	-0.0031	0.0034
x ₁ y ₃	0.0189	0.0017	0.0191	0.0015
x ₁ y ₄	0.0047	0.0040	0.0054	0.0032
x ₁ x ₁	-0.0251	0.0087	-0.0216	0.0068
x ₂ y ₁	-0.0802	0.0126	-0.0899	0.0080
x ₂ y ₂	-0.0202	0.0121	-0.0141	0.0060
x ₂ y ₃	-0.0107	0.0033	-0.0101	0.0028
x ₂ y ₄	0.0542	0.0092	0.0553	0.0053
x ₂ x ₁	0.0020	0.0150	0.0066	0.0108
x ₂ x ₂	-0.3136	0.0355	-0.2655	0.0203
x ₃ y ₁	0.0312	0.0150	0.0387	0.0099
x ₃ y ₂	-0.0204	0.0112	-0.0195	0.0071
x ₃ y ₃	0.0064	0.0043	0.0056	0.0035
x ₃ y ₄	-0.0465	0.0100	-0.0466	0.0064
x ₃ x ₁	0.1143	0.0150	0.1111	0.0107
x ₃ x ₂	0.1382	0.0282	0.1114	0.0173
x ₃ x ₃	0.2650	0.0445	0.2560	0.0280
x ₄ y ₁	-0.0325	0.0110	-0.0318	0.0079
x ₄ y ₂	0.0274	0.0077	0.0262	0.0048
x ₄ y ₃	-0.0161	0.0033	-0.0160	0.0027
x ₄ y ₄	-0.0006	0.0075	-0.0017	0.0052
x ₄ x ₁	-0.0511	0.0107	-0.0514	0.0080
x ₄ x ₂	0.0372	0.0217	0.0447	0.0138
x ₄ x ₃	-0.2723	0.0278	-0.2694	0.0169

x ₄ x ₄	0.1292	0.0229	0.1246	0.0147
x ₅ y ₁	0.0457	0.0061	0.0463	0.0046
x ₅ y ₂	0.0228	0.0051	0.0210	0.0036
x ₅ y ₃	0.0043	0.0016	0.0039	0.0014
x ₅ y ₄	-0.0167	0.0046	-0.0162	0.0036
x ₅ x ₁	-0.0188	0.0073	-0.0201	0.0056
x ₅ x ₂	0.0836	0.0131	0.0759	0.0093
x ₅ x ₃	-0.1433	0.0112	-0.1293	0.0084
x ₅ x ₄	0.0716	0.0103	0.0699	0.0078
x ₅ x ₅	-0.0094	0.0079	-0.0116	0.0064
y ₁ t	-0.0002	0.0003	-0.0003	0.0002
y ₂ t	0.0003	0.0003	0.0003	0.0001
y ₃ t	0.0004	0.0001	0.0004	0.0001
y ₄ t	0.0006	0.0002	0.0006	0.0001
x ₁ t	0.0010	0.0004	0.0009	0.0002
x ₂ t	-0.0040	0.0007	-0.0034	0.0003
x ₃ t	-0.0016	0.0008	-0.0016	0.0004
x ₄ t	0.0028	0.0006	0.0027	0.0003
x ₄ t	0.0022	0.0003	0.0021	0.0002
t	-	-	0.0083	0.0016

Note: $y_1 = \log(\text{commercial \& industrial loans/real estate loans})$, $y_2 = \log(\text{loans to individuals/real estate loans})$, $y_3 = \log(\text{securities/real estate loans})$, $y_4 = \log(\text{off-balance sheet items/real estate loans})$, $x_1 = \log(\text{total demand deposits})$, $x_2 = \log(\text{total time \& savings deposits})$, $x_3 = \log(\text{labor})$, $x_4 = \log(\text{capital})$, $x_5 = \log(\text{purchased funds})$ and $t = \text{time trend}$.

Figure 1: Average Efficiencies over Time, the Spline Specification

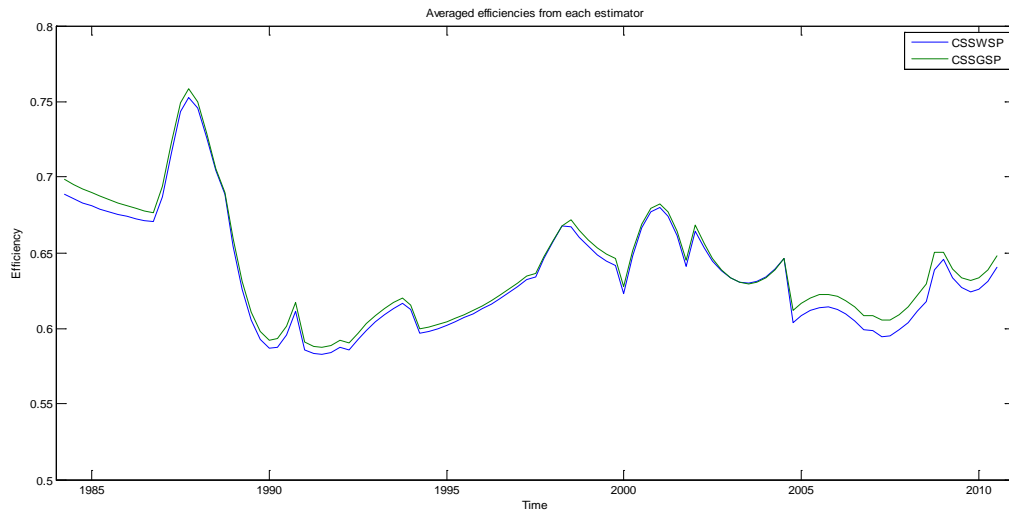


Figure 2: Average Efficiencies over Time, the Polynomial Specification

