On the Estimation of Marginal Cost

Manthos D. Delis, Maria Iosifidi

Surrey Business School, University of Surrey, Guildford, GU2 7XH, UK,
m.delis@surrey.ac.uk, m.iosifidi@surrey.ac.uk

Efthymios G. Tsionas

Department of Economics, Athens University of Economics and Business
76 Patission Street, 10434, Athens, Greece, tsionas@aueb.gr

This article proposes the estimation of marginal cost of individual firms using semiparametric and nonparametric methods. These methods have a number of appealing features when applied to cost functions. The empirical analysis uses data from a unique sample of the California electricity industry for which we observe the actual marginal cost and estimate the marginal cost from these data. We compare the actual values of marginal cost with the estimates from semiparametric and nonparametric methods, as well as with the estimates obtained through conventional parametric methods. We show that the semiparametric and nonparametric methods produce marginal cost estimates that very closely approximate the actual. In contrast, the results from conventional parametric methods are significantly biased and provide invalid inference.

Key words: estimation of marginal cost; parametric models; semi-parametric models; nonparametric models; California electricity industry; actual and simulated data

1. Introduction

Since the contribution of Alfred Marshall, one of the most fundamental and widely used concepts in economics and management sciences is marginal cost. However, in most industries, the marginal cost of firms cannot be readily observed in the data. Thus, researchers and practitioners have to rely on estimates that might not be robust. This problem creates notorious difficulties in empirically analyzing basic microeconomic theories on industrial organization and management. These difficulties also extend to policy makers and their evaluation of the cost structures of firms and industry conduct. This paper revisits the issue of the estimation of marginal cost with a standard cost function and shows how to derive robust estimates for individual firms in our sample and for each point in time (i.e., at the observation level).

In particular, we propose the estimation of a standard cost function that uses a class of semiparametric or nonparametric estimation methods; namely, the smooth coefficient model of Fan (1992), the generalized additive model of Hastie and Tibshirani (1986), and the nonparametric inefficiency model of Kumbhakar et al. (2007). These flexible methods allow the relaxation of a number of restrictive assumptions that pertain to the estimation of cost functions; the most important being the assumption of a specific functional form that is required under parametric methods. Reiss and Wolak (2007), among many others, are skeptical about this assumption, because the structure of the cost and output data can bias marginal cost estimates to an unknown magnitude and direction. In contrast, under the semiparametric and nonparametric methods, variables can vary only according to information derived from the data.

Even though the semiparametric and nonparametric methods are theoretically well-established in the statistics literature, researchers have not verified using actual data that the true values are equal or close to the estimated values of the parameters. Because of the lack of studies on this area, researchers are sceptical about the applicability of this method to estimate marginal cost. To confirm that the estimated marginal cost from the nonparametric methods closely approximates the true marginal cost, we conduct a number of empirical tests. Specifically, we use data and implications from important studies on one of the few industries where the true values of marginal cost can be readily observed from the available data and estimated from the same data set: California's electricity industry. We then compare the values of the true marginal cost to these of its estimates and show that the correlation between the two is very high and that their distribution densities are quite similar. We carry out these tests using two data sets: actual aggregate data for all firms in the industry and simulated panel data that comprise information for each firm at each point in time.

In addition, we estimate the same total-cost equation using state-of-the-art parametric models. In particular, we use the two data sets and impose various parametric forms to the cost function. We find that marginal cost estimates using parametric methods are relatively worse approximations of true marginal costs as compared to the nonparametric methods. For the simulated panel data set where different forms have different production technologies, the bias is so large that marginal cost estimates based on parametric methods result in invalid inference.

Empirical studies of marginal cost date back at least to Rosse (1967). In most of these studies, researchers carry out marginal cost estimations by using parametric econometric methods, with the assumption that production technology is constant among firms in the same industry or between different groups of firms. Therefore, based on our empirical analysis, the results of these studies might be considerably biased. The realization of this problem, among other issues, leads some researchers to avoid estimating cost equations and to rely on alternative empirical techniques to infer firm behavior and industry conduct (e.g., by using demand equations, as in Nevo, 2001). Other important contributions use data from the few industries where researchers can observe the true marginal cost. For example, Wolfram (1999) and Borenstein et al. (2002) use such data from the electricity industry and Genesove and Mullin (1998) from the sugar industry.

The rest of the paper proceeds as follows. Section 2 discusses the data set of Borenstein et al. (2002) from California's electricity market and presents the results from the parametric methods. Section 3 re-estimates marginal costs by using three semiparametric and nonparametric methods. Section 4 carries out the analysis of the two previous sections based on simulated data. Section 5 concludes.

2. Data and Estimation of Marginal Cost using Parametric Methods

2.1. Data from California's Electricity Market

This section describes the actual data set from California's electricity industry. Borenstein et al. (2002) use hourly data of electricity generation firms operating in the California electricity market over the period June 1998 to October 2000. Since these data only provide the required information at the aggregate level across all firms in the industry for each hour, they are time-series data. We have information on the actual marginal cost, as well as on the data required to estimate the cost function and derive the marginal cost. Thus, we can compare the actual marginal cost with its estimates through the use of basic statistical methods to infer whether the estimates closely approximate the true marginal cost.

The data set comprises 21,217 observations on marginal cost (mc), total cost (tc), industry output (q) measured in MW per hour, and the input prices of production measured in \$/MWh. Most of the

firms in the industry use two input prices, namely the average daily price of natural gas for California (png) and the price of NOx permits (pno). These are the input prices that Kim and Knittel (2006) also use when they employ the same data set to estimate the cost function with parametric methods. We report summary statistics for these variables in Table 1.

[INSERT TABLE 1]

Figures 1 to 3 offer a graphical representation of the relationship between total cost and (i) electricity output, (ii) the price of natural gas, and (iii) the price of NOx permits. In all three figures, we add a locally weighted regression line to provide a first analysis of the determinants of total cost. Figure 1 shows that the relationship between total cost and output is somewhat different between two classes of production units. The upper part of the figure shows observations that reflect a convex function, whereas those in the lower part of the figure reflect an almost linear relationship. Figure 2 shows that the relationship between total cost and the price of natural gas can be characterized by a convex function, while Figure 3 shows that the equivalent relationship between total cost and the price of NOx permits is represented by a concave function. The large variation in the prices of natural gas and NOx permits stems from the fact that during the California electricity crisis that began early in 2000, input prices increased drastically, becoming a significant component of changes in costs. This crisis is essentially the reason for the observed non-linearity in Figures 1 to 3.

[INSERT FIGURES 1 to 3]

2.2. Estimation of Marginal Cost using Parametric Methods

Before discussing the semiparametric and nonparametric methods to estimate the cost equation, we use existing parametric methods and the implications of the extensive literature on this issue to obtain marginal cost estimates. There are several studies with various objectives that estimate the marginal cost of firms. Only indicatively, see Hall (1988), Bresnahan (1989), Roeger (1995), Konings et al. (2005), and Koetter et al. (2012). A common feature of these studies is that they rely on strong assumptions of the functional form, usually without an initial analysis on the properties of their data. Indeed, most of these studies employ a translog functional form without any prior tests. We show that

such an *ad hoc* analysis using parametric methods can lead to erroneous estimation of marginal cost. This observation raises serious doubts about the findings and implications of the literature on several microeconomic, industrial-organization, and management-science projects.

We closely follow the conventional approaches to estimate marginal cost (e.g., Koetter et al., 2012). We specify various parametric forms for the cost function, such as the log-linear and the more flexible translog, generalized Leontief, and Fourier. We present the results from the log-linear and the translog cost functions, those that are used most commonly in the literature. The log-linear cost function (log of Cobb-Douglas) takes the following form:

$$\ln(tc / pno) = c_0 + c_1 \ln(png / pno) + c_2 \ln q + u, \quad (1)$$

where tc is total cost, q is industry output, png is the average daily price of natural gas for California and pno is the price of NOx permits. The translog cost function takes the following form:

$$\begin{aligned} \ln(tc / pno) = c_0 + c_1 \ln(png / pno) + c_2 \ln q + c_3 \ln(png / pno) * \ln q + \\ c_4 1/2 * (\ln(png / pno))^2 + c_5 1/2 * (\ln q)^2 + u. \end{aligned} \quad (2)$$

In these specifications, we impose homogeneity of degree one on input prices by dividing png and tc by pno . From (1) and (2), respectively, marginal cost can be calculated for each observation as follows:

$$mc = tc/q * c_2 \quad (3)$$

$$mc = tc / q [c_2 + c_3 \ln(png / pno) + c_5 \ln q]. \quad (4)$$

We begin by estimating (1) and (2) with OLS. As the values of the variables added to these equations vary, this method yields observation-specific estimates of mc for both (1) and (2). Evidently, this is a crucial issue underlying the shape of the values of the estimated mc , as this depends not only on the coefficient estimates but also on the variables themselves. We report the coefficient estimates and t -statistics in the first two columns of Table 2. All estimated coefficients are statistically significant at the 1% level and bear the expected sign. Also, high adjusted R-squared values show that very little is left unexplained.

[INSERT TABLE 2]

With these specifications, we assume that output is exogenous to the cost equation. Reiss and Wolak (2007) note that the exogeneity of output and input prices in cost functions depends on the type of the

industry examined. Specifically, endogeneity of q arises if, for example, the cost function comes from a production function $q_i = f(A_i, L_i, K_i)$ that includes component A_i characterizing technology as different among firms. Then, the cost equation will also be a function of A and if A is unobserved it renders q endogenous. In this case there are two choices: to identify reasonable estimates of A_i and add them to the estimated equation and to instrument q . In our setting, the extent to which the estimated marginal cost approximates the true marginal cost can guide us through this potential identification issue. Furthermore, the technology used by different firms in the electricity industry should not be widely different, and instrumental variables' regressions should be less efficient than ordinary least squares (OLS) if the endogeneity bias is small. The third and fourth columns of Table 2 report the results from the estimations of (2) and (3) using a two-stage least squares (2SLS) estimation method and the changes in the two input prices as instruments. The California electricity firms consider input prices as exogenous (Borenstein et al., 2002), and therefore their changes qualify as proper instrumental variables.

Using (3) and (4), we obtain the marginal cost estimates for the four different specifications in Table 2. We report summary statistics for these estimates in Panel A of Table 3, which show that the mean values from the different specifications closely approximate the mean value of the true marginal cost. In this respect, the parametric models approximate the true marginal cost fairly well. Next, we examine the correlation coefficients between the parametric estimates and the true values of marginal cost (see Panel B of Table 3). The higher correlation is between the marginal cost obtained from the estimation of the translog with OLS (mc_{tols}) and the true marginal cost (mc) and is equal to 0.902. The rest of the relevant correlation coefficients in columns (1) to (4) are below 0.9.

[INSERT TABLES 3]

However, more importantly, comparing the probability density function (PDF) of marginal cost estimates with the equivalent of the true marginal cost (shown in Figures 4 to 7) divulges important differences. The first two figures show that the estimation of the cost function with OLS somewhat overestimates the total cost for a significant number of observations. The t-test and sign-test for the equality of values (reported below the figures) strongly reject the null hypothesis of the equality of the pairs of series (for an explicit description of these tests, see Hoel 1966). Further, the results from both

the log-linear and the translog specifications illustrate that 2SLS produces marginal cost estimates that deviate from true marginal cost to a larger degree. Specifically, the correlation coefficients between mc and mc_{ll2sls} or mc_{tr2sls} are lower compared with their equivalents from the OLS. Most importantly, the PDFs of mc_{ll2sls} and mc_{tr2sls} show a worse fit to the true marginal cost compared to the OLS equivalents (see Figures 6 and 7 and associated t-tests and sign-tests). Therefore, this is evidence that any use of these estimates for research or policy purposes biases the inference, even for an industry with simple technology such as the one described by this data set.

[INSERT FIGURES 4 to 8]

So far we have used the main functional forms employed in the related literature. Despite the fact that the majority of this literature does not test for the suitability of the functional form, we find that both the log-linear and the translog models do not pass the Ramsey (1969) test for neglected non-linearity in the choice of functional form (p-values in both cases equal 0.000). Given the pictures presented in Figures 1 to 3, we use a specification of the form

$$\ln(tc/pno) = c_0 + c_1 \ln(png/pno) + c_2 \ln q + c_3 (\ln(png/pno))^2 + c_4 (\ln q)^2, \quad (5)$$

which simply adds the squared terms of the output and the scaled input price as determinants of total cost. The p-value of the Ramsey test now increases to 0.02, still rejecting the null at the 5% level. We present the PDF from this functional form in Figure 8. The fit is relatively better; however, the t-test for equality is still statistically significant at the 1% level and the sign-test rejects the null of equality of distributions of the estimates with the true marginal cost. We experiment with other functional forms such as the generalized Leontief and the Fourier, or with *ad hoc* functional forms that include various combinations of squared terms, but we always reject the null hypothesis for the suitability of the functional form.

Furthermore, we conduct sensitivity analyses on these findings (i) by estimating (1), (2), and (5) using the maximum likelihood method, and (ii) by removing the outliers of marginal cost (we drop 0.5% of the sample from each end of the distribution of marginal cost). The results from these sensitivity analyses do not improve the results already presented. We feel that the main driver of these findings is the presence of the variables tc/q , $\ln(png/pno)$ and $\ln q$ in (3) and / or (4), which is imposed by the

parametric assumptions. Indeed, one can experiment with other variants of parametric models and perhaps find a specification that yields estimates that more closely approximate the true values of marginal cost for the present sample. However, in most industries, the researcher or the practitioner is not aware of the true marginal cost, making comparisons with estimates difficult; this is why researchers work with estimation results.

3. Estimation with Semiparametric and Nonparametric Methods

This section presents the results for the estimation of marginal cost using the data from the California electricity market and three semiparametric and nonparametric methods. These methods are quite general and can be applied to any industry in which the true marginal cost is not observed. The reason for this is that these methods do not rely on any specific parametric assumptions about the type of technology used in the industry or by different firms within the industry and the extent of the market. The only requirements are as follows: (i) the econometric methodology must be robust to a reasonable degree and (ii) the variables employed should not have measurement error.

Given these properties, we rely on the estimation of a general linear cost function in the form of

$$tc_{it} = a_0 + a_1 png_{it} + a_2 pno_{it} + a_3 q_{it} + e_{it} . \quad (6)$$

For this cost equation, marginal cost is simply equal to a_3 , and hence estimation is not affected by the structure of the data (i.e., no actual variables are involved in the calculation of marginal cost after estimation). Note that all of the estimated models are varying-coefficient models because they allow as many estimates for a_3 as the observations in the data.

3.1. Semiparametric and Nonparametric Methods

We utilize the semiparametric smooth coefficient model of Fan (1992), the generalized additive model of Hastie and Tibshirani (1986), and the nonparametric inefficiency model of Kumbhakar et al. (2007). Due to space constraints, we provide only few details on the estimation of the three models using the data on California's electricity industry.

We first discuss the smooth coefficient model. Our initial assumption is that the data are $\{Y_i, W_i\}$, $i = 1, \dots, n$, where Y is the response variable that equals tc , and W is the matrix of the predictor variables that comprises png , pno , and q . We suppress the subscript t that represents hours. Also, we have $W_i = \{X_i, V_i\}$, where X comprises the input prices png and pno , while V is the non-parametric part of the model (in our case just q). We can now rewrite (6) in econometric form as follows:

$$Y_i = E(Y_i|W_i) + e_i = X_i\beta_1 + V_i\beta_2(z_i) + e_i. \quad (7)$$

In (7), β_2 is a function of one or more variables with dimension k added to the vector Z , which is an important element of the analysis and will be discussed below. The linear part in (7) is in line with the idea of the semiparametric model as opposed to a nonparametric model (e.g., Zhang et al., 2002). The coefficients of the linear part are estimated in the first step as averages of the polynomial fitting by using an initial bandwidth chosen by cross-validation (Hoover et al., 1998). We then average these estimates β_{1i} and β_{2i} to receive β_1 and β_2 in (7). In the second step we use these average estimates and (7) to redefine the dependent variable as follows:

$$Y_i^* = V_i\beta_2(z_i) + e_i^*, \quad (8)$$

where the asterisks denote the redefined dependent variable and error term.

$\beta_2(z)$ is a vector of smooth but unknown functions of z_i , estimated using a local least squares approach of the form

$$\begin{aligned} \hat{\beta}_2(z) &= \left[(n\lambda^k)^{-1} \sum_{j=1}^n V_j^2 K\left(\frac{z_j - z}{\lambda}\right) \right]^{-1} \left[(n\lambda^k)^{-1} \sum_{j=1}^n V_j Y_j^* K\left(\frac{z_j - z}{\lambda}\right) \right] \\ &= [B_n(z)]^{-1} C_n(z), \end{aligned} \quad (9)$$

where $B_n(z) = (n\lambda^k)^{-1} \sum_{j=1}^n V_j^2 K\left(\frac{z_j - z}{\lambda}\right)$, $C_n(z) = (n\lambda^k)^{-1} \sum_{j=1}^n V_j Y_j^* K\left(\frac{z_j - z}{\lambda}\right)$.

Equation (9) represents a local constant estimator, where $K(z, \lambda)$ is a kernel function, λ is the smoothing parameter (chosen by generalized cross validation in all models of our paper) for sample size n , and k is the dimension of z_i .

If we assume that z is a scalar and K is a uniform kernel, then (9) can be written as follows:

$$\hat{\beta}_2(z) = \left[\sum_{|z_j - z| \leq \lambda} V_j^2 \right]^{-1} \left[\sum_{|z_j - z| \leq \lambda} V_j Y_j^* \right]. \quad (10)$$

In (10), $\hat{\beta}_2(z)$ is a least squares estimator obtained by regressing Y_j^* on V_j , using the observations of (V_j, Y_j^*) for which the corresponding z_j is close to z , that is, $|z_j - z| \leq \lambda$. Therefore, to estimate $\hat{\beta}_2(z)$, we only use observations within this “sliding window.” Note that no assumptions are made about this estimator globally, but locally—within the sliding window—we assume that $\hat{\beta}_2(z)$ can be well-approximated. Also, because $\beta_2(z)$ is a smooth function of z , $|\beta_2(z_j) - \beta_2(z)|$ is small when $|z_j - z|$ is small. The condition that $n\lambda$ is large ensures that we have sufficient observations within the interval $|z_j - z| \leq \lambda$ when $\beta_2(z_j)$ is close to $\beta_2(z)$. Therefore, under the conditions $\lambda \rightarrow 0$ and $n\lambda^k \rightarrow \infty$ (for $k \geq 1$), the local least squares regression of Y_j^* on V_j provides a consistent estimate of $\beta_2(z)$ (for a proof, see Li et al., 2002). Therefore, the estimation method is usually referred to as a local regression.

A critical issue in the estimation process is the choice of the variable(s) to comprise z . The best candidates are variables highly correlated with β_2 , but that also allow variation for β_2 across firms and time. In a cost function, the natural candidates to use are the input prices as z . The advantage of this choice is that input prices most certainly affect β_2 to a large extent. This has been shown many times when researchers employ a translog specification, which includes multiplicative terms of output with input prices, to estimate the cost function parametrically. In our sample, specification (2) of Table 2 provides evidence that this is indeed the case. Thus, we employ $(png + pno)/2$ as z , and as a robustness check we also experiment with both the two input prices separately and find no significant changes in the results.

The second model we use in this study is the generalized additive model of Hastie and Tibshirani (1986). Wood (2006) has a thorough discussion of this model, thus we only present here the practical issues for our estimations. The link function $g(\mu)$ for our three predictors takes the form

$$g(\mu) = a + \sum_{j=1}^{q=3} f_j(W_j), \quad (11)$$

where f are smooth functions of the three covariates and q is the basis dimension that we discuss below. The “additive” term refers to the fact that the covariates are assumed to be additive on the scale of the link function. The “generalized” term alludes to the fact that we can use the error structures and link functions that are available with generalized linear models. This model is fit with the penalized local likelihood procedure described by Hastie and Tibshirani (1986) and Wood (2006).

To estimate the generalized additive model we need to make some assumptions. First, to estimate f , (11) needs to be represented in such a way that it becomes a linear model. This can be done by choosing a basis, which defines the space of functions of which f (or a close approximation of it) is an element. Choosing a basis is equivalent to choosing some function which will be treated as completely known. For example, if $b_i(W)$ is the i^{th} such basis function, then f will have the representation $f(W) = \sum_{i=1}^q b_i(W)\beta_i$, for some values of the unknown parameters β_i . Substituting this expression in (11) yields a linear model, which can be estimated with least squares (Wood, 2006). We choose the cubic smoothing spline, which is easy to compute and allows the program to run for a big sample size, such as ours (also see Wood, 2006). This consists of sections of a cubic polynomial, joined together so that they form a continuous function. The points at which these sections are joined are known as knots of the spline, which define the basis dimension q . As Wood (2006) suggests, the exact size of the basis dimension is not that critical, as this dimension is only setting an upper bound of flexibility of the function. Hence, the model’s fit is more or less insensitive to the basis dimension, as long as this provides enough flexibility. We experiment with various basis dimensions and report the results from $q = 4$, as our results become more sensitive for lower values of q .

The third model employed is the inefficiency model by Kumbhakar et al. (2007) and Kumbhakar and Tsionas (2008). These methods utilize a model of the form,

$$Y = f(z) - u + e, \quad (12)$$

which is a localized version of the equivalent parametric model of Aigner et al. (1977). In this model, z are the predictors, $u > 0$ is the inefficiency term, and e represents random noise. Under the “initial” assumption of linearity and normal errors terms, we have

$$Y_i = z_i' \alpha_i - u_i + e_i, \quad u_i \sim iidN(0, \sigma_{ui}^2), \quad e_i \sim iidN(0, \sigma_{ei}^2). \quad (13)$$

Here we assume that $a_i = a(z_i)$, $\sigma_{ui}^2 = \exp[\log \sigma_u^2(z_i)]$, $\sigma_{ei}^2 = \exp[\log \sigma_e^2(z_i)]$, which we denote collectively by $\theta_i = \theta(z_i)$. The specification is quite general, as all the coefficients and variances are permitted to vary among observations.

The corresponding likelihood for the i^{th} observation is as follows:

$$l_i(Y_i, \theta(z_i)) = (2\pi\sigma^2(z_i))^{-1/2} \exp\left[-\frac{(Y_i - z_i'\alpha(z_i))^2}{2\sigma^2(z_i)}\right] \cdot \Phi\left(-\frac{\lambda(z_i) \cdot (Y_i - z_i'\alpha(z_i))}{\sigma(z_i)}\right). \quad (14)$$

In (14) Φ denotes the standard normal distribution function, $\sigma^2(z_i) = \sigma_e^2(z_i) + \sigma_u^2(z_i)$, and

$\lambda(z_i) = \frac{\sigma_u(z_i)}{\sigma_e(z_i)}$. The global log-likelihood function is as follows:

$$L(Y, \theta) = \sum_{i=1}^n \log l_i(Y_i, \theta(z_i)) w(z_i), \quad (15)$$

where $w(z_i)$ denotes the weights. Here we use the local quadratic approximation following Kumbhakar and Tsionas (2008):

$$\theta(z) = \langle a_o, A(z_i - z) \rangle = a_{oo} + a_o'(z_i - z) + \frac{1}{2}(z_i - z)' A(z_i - z), \quad (16)$$

where $\langle \cdot \rangle$ denotes the inner product and a_o, A are a vector and a symmetric matrix, respectively, of unknown coefficients. Suppose $a = [a_o', \text{vech}(A)']'$. This means that we use the local quadratic approximation for both the coefficients and the log variance $\log \sigma^2(z_i)$. We use the log-variance specification following Kumbhakar et al. (2007) to keep the variances positive. Relative to local linear estimation this log-likelihood specification is more general in the sense that (i) we allow for arbitrary form of heteroskedasticity and (ii) we do not effectively rely on the normality assumption. The local likelihood estimate of $\theta(z)$ is $\hat{\theta}(z) = \hat{a}_{oo}$. The weight function is as described before and the bandwidth parameters are determined using generalized cross-validation.

There are three main advantages of the latter approach. First, we do not assume that variances are constant and at the same time we model non-parametrically the distribution of the error term. Second, we model the inefficiency component of the error term, which might be important in empirical applications involving the estimation of cost functions. Furthermore, relative to Hastie and Tibshirani

(1986), the efficiency model does not use an arbitrary local span or “window” of observations to yield smooth estimates, but rather a multivariate kernel. Setting a window would make the results sensitive in practice, whereas localization using the full likelihood based on cross-validation is fully automatic and does not require arbitrary choices of windows. Similar experience has been found and reported in Kumbhakar et al. (2007). The multivariate kernel has well-established asymptotic properties, which can serve as a rough guidance to the good behavior of the estimator in small samples as reported again in the simulation experiments of Kumbhakar et al. (2007). Even though this leads to increased computational cost, this model is useful in practice and is likely to yield better results particularly in finite samples.

3.2. Empirical Findings

This section reports and analyzes the estimation results from the smooth coefficient model, the generalized additive model, and the inefficiency model, and compares these results with the true marginal cost that the analysis in Section 2 provides. We demonstrate that the semiparametric and nonparametric estimates of marginal cost closely approximate the true marginal cost, which shows that the proposed methods provide improved estimates of marginal cost compared with their equivalents from the parametric methods.

Panel A of Table 4 reports summary statistics for the coefficient estimates of png , pno , and q from the smooth coefficient model. The average of the estimated coefficients for q — the marginal cost (denoted as mc_{scm}) — is 35.10 and is very close to the average of the true marginal cost (34.45). Similarly, the correlation coefficient between mc_{scm} and the true mc , reported in the last row of Panel A, is as high as 0.974. Furthermore, Figure 9 depicts the graph of the PDF of mc_{scm} against that of the true mc as done for the parametric models earlier. The figure shows a relatively better fit than before with the estimated marginal cost mapping the true marginal cost quite well. Also, the p-values of the t-test and the sign-test do not reject the null hypothesis that the two PDFs are equal. This is first-hand evidence that the proposed method provides superior estimates of the marginal cost at each point in the data compared with the estimates from the parametric models.

[INSERT TABLE 4]

[INSERT FIGURES 9–11]

Panel B of Table 4 and Figure 10 report the results from the generalized additive model. These results represent a marginal improvement over the equivalent ones from the smooth coefficient model. The correlation coefficient between the marginal cost from the generalized additive model (mc_{gam}) and mc is marginally higher (0.983). Furthermore, the PDF of mc_{gam} is almost identical to that of mc , while the associated t-test and sign-test yield higher p-values compared with their equivalents from the smooth coefficient model. Thus, we can conclude that the generalized additive model is particularly suitable for the estimation of marginal cost in the California electricity market.

The results from the generalized additive model are also very close to those from the nonparametric inefficiency model. Panel C of Table 4 shows that the correlation coefficient between the marginal cost from the inefficiency model (denoted as mc_{im}) and mc is as high as 0.987. This is the highest correlation coefficient between an estimated marginal cost and the true marginal cost identified among all of our estimations. The p-values of the t-test of equality of the two series and the sign-test of the equality of the two PDFs are also very high, confirming a very good fit of mc_{im} to mc (see Figure 11). Another interesting result from the inefficiency model is the estimation of the inefficiency component u . These estimates show that the inefficiency is not particularly large in our sample, with the average value of u being 0.07, the minimum value being zero by construction, and the maximum value being 0.13.

We conduct a number of sensitivity analyses on the results of this section. In particular, (i) we use only png as Z in the smooth coefficient model, instead of the linear combination of both input prices, (ii) we include a time trend among the regressors, and (iii) we use a number of different options for the degrees of freedom for the predictors, the distribution of the dependent variable, etc. The results, unreported here but available on request, are not significantly different from those reported in Table 4 and shown in Figures 9 to 11.

Our results show that the semiparametric and the nonparametric methods for the estimation of marginal cost represent a great improvement over the parametric estimation methods. The specification tests in Table 4 are quite definite in that the inference based on the estimates from the semiparametric and non-parametric methods can be carried out with confidence. For example, if one has information

on prices of products and needs marginal cost estimates to calculate price-cost margins, then the estimates from the proposed methods will approximate the true price-cost margins relatively better than the estimates from the conventional methods. Many other examples can be provided from the economics and business literatures where estimates of marginal cost are required for research and policy-making reasons. Below we confirm that these conclusions also hold for a simulated panel data set.

4. Evidence from Simulated Panel Data

Most research studies that estimate marginal cost from the cost function use panel data from firms operating in a particular industry. The difference between studies that use panel data and those that use aggregate data can be considerable if firms have different production technologies and cost structures. This difference implies that imposing parametric structures on the cost function can yield larger biases in panel data studies. To examine the hypothesis that the true marginal cost equals the estimated one using panel data, we generate simulated-panel data from the underlying information that the actual data provides.

In particular, we conduct a data generation process for the variables q (now output of firm i), png , pno , and mc . The dimension of the panel is set to 4,000 hourly observations each for five firms that total to 20,000 observations. We also experiment by generating 21,217 observations each for five firms, i.e., 106,085 observations in total. However, the larger number of observations in this panel increases the computational burden of the semiparametric and nonparametric analyses in the software used for the estimation (Gauss, Stata, and R), most of the times rendering the estimation impossible on a CORE i7 processor with 6GB of RAM.

To generate data for q , png , and pno , we use a chi-squared distribution and the mean values from the actual data set (i.e., 5,006.8, 27.30, and 3.87, respectively). We favor the chi-squared distribution over, for example, the log-normal distribution because PDFs of the actual data for these variables are positively skewed. We define q as exogenous to the cost equation, which has important implications for the validity of the estimation procedure below, as one can use OLS- and likelihood-based methods.

Finally, to generate data on mc we follow the paradigm of Kim and Knittel (2006). Specifically, we estimate, using the actual data and OLS, a linear regression of the form

$$mc_{it} = f(png_{it}, pno_{it}, q_{it}, \sum_1^3 d), \quad (17)$$

where d contains the parameters on the independent variables. This yields

$$mc = 0.570png + 0.750pno + 0.002q + 4.253 + \varepsilon, \quad (18)$$

which are the coefficients used for the data generation process for mc . Given the positive skewness of mc in the actual data, in (18) the stochastic term ε is chi-squared distributed with the mean of the actual dataset (34.45). For the generation of mc , we also experiment with all other functional forms used to estimate the parametric models in Section 2.2. Not surprisingly, in the case for example of the log-linear model (Equation 3) the best fit to the simulated mc are the estimates of mc from the log-linear specification, while in the case of the translog model (Equation 4) the best fit are the estimates of mc from the translog specification. Consequently, these tests are not very informative if one wants to examine cases where the underlying structure of the data is more complex.

We provide summary statistics for the generated variables in Table 5. From this data set, we calculate total costs for each available observation from the textbook formula:

$$mc = \Delta tc / \Delta q \Rightarrow mc = (tc_1 - tc_0) / (q_1 - q_0) \Rightarrow tc_1 = mc(q_1 - q_0) + tc_0. \quad (19)$$

This is a good approximation for tc , because of the high frequency of the data. This data generation process yields a tc that has a bimodal distribution, which poses a more stringent requirement on the parametric and nonparametric models to be studied later. Thus, the requirement is that the estimates of marginal cost must be able to approximate the true marginal cost even though the total cost is clustered in two modes. This clustering could arise, for example, from an industry where there are two groups of firms such as a number of large firms with high costs and a number of small ones with lower costs, or from an industry where there is seasonality in the cost data with a period of high costs and a period of low costs.

Table 6 reports the results from applying OLS with fixed and time effects to the log-linear and the translog specifications. Similar to our analysis in Section 2.2, other functional forms for the parametric specification of the cost equation do not substantially improve the fit of the marginal cost estimates to

the actual values of the marginal cost, and are not reported due to space constraints. Both the coefficients on the price and the quantity variables added to the log-linear specification and the “actual” effect of these variables in the translog specification (when taking partial effects) are positive and statistically significant at the 1% level. We proceed by estimating mc from each specification, given (3) and (4). The mean values of the estimated marginal cost, reported in Panel A of Table 7, are higher than the those of the actual marginal cost reported in Table 5. This shows that, on average, both parametric specifications overestimate marginal cost in the sample. More importantly, the correlation coefficient between the estimates of marginal cost obtained from the log-linear specification (mc_{llols}) and the actual values of marginal cost is nearly zero (see Panel B of Table 7). The equivalent correlation coefficient involving the estimates from the translog is better, but still as low as 0.342.

[INSERT TABLES 5 to 8]

The reason behind these findings is the fact that the PDF of tc is bimodal. Figure 12 presents the PDF of mc_{llols} versus that of the actual mc . Even though the PDF of tc is bimodal, the PDF of the true $mc = \Delta tc / \Delta q$ is unimodal. In contrast, the PDF of mc_{llols} is bimodal, which is imposed by the appearance of tc in (3) and no other terms that improve the flexibility of the estimates. Therefore, the estimates of marginal cost from the log-linear specification can be severely biased if the structure of the cost data is unfavorable. As shown in Figure 13, the estimates’ flexibility is significantly improved when using the translog. Yet, the fit of the estimates to the actual mc is still far from optimal, with the estimates of mc displaying a relatively higher standard deviation (see also relevant values in Tables 5 and 7). Also, the p-values of the t-test and the sign-test are essentially equal to zero, showing that the estimated and the true marginal costs are far from equal.

[INSERT FIGURES 12 to 16]

We now turn to the estimation of the cost equation, using the semiparametric and nonparametric methods. We report averages of the coefficient estimates from the smooth coefficient model in Panel A of Table 8. Again, the mean of the estimates on mc_{scm} is close to that of the true mc from the simulated panel (34.7 versus 34.4). Most importantly, the correlation coefficient between the two is as high as 95%, which is a great improvement compared to the equivalent one from the translog equation of 34%. Further, we compare the PDF of mc_{scm} obtained from the simulated data set to the true mc in Figure 14.

Again, the mapping of the values of mc_{scm} against those of the true mc improves significantly compared to those in Figures 12 and 13. Also, the p-values of the t-test and the sign-test are higher, even though the sign-test rejects the null hypothesis of the equality of the two PDFs at the 5% level of significance.

Therefore, the most suitable method for the estimation of marginal cost using the simulated data seems to be the generalized additive model. We report the results from this method in Panel B of Table 8 and compare the PDFs of the estimated and the actual marginal costs in Figure 15. The correlation coefficient between mc_{gam} and mc is as high as 0.966 and the p-values of the t-test and the sign test suggest that the null hypothesis of equality is not rejected. This considerable improvement shows that if the marginal cost estimates from the generalized additive model are used to, for example, calculate price-cost margins, then the accuracy of the margins will improve considerably.

Finally, Panel C of Table 8 and Figure 16 present the results from the inefficiency model. The inefficiency model produces estimates that are relatively better than the estimates from the smooth coefficient model and marginally inferior to those of the generalized additive model. The p-values of the t-test and the sign-test show that the respective null hypotheses of equality are not rejected at the 5% level (marginally for the sign-test). Furthermore, similar to the case of the actual data, the inefficiency model predicts a relatively low inefficiency, which is on average equal to 8%. The marginal superiority of the generalized additive and the inefficiency models over the smooth coefficient model might stem from the fact that they are not prone to the curse of dimensionality. Given this feature, the two nonparametric models should also be applicable in cases with smaller, less dense data sets. It remains to be examined whether the inefficiency model fits the data better than the generalized additive model when the firms in the industry operate under higher inefficiency levels. We leave this exercise for future research.

5. Conclusions

Research and policy have long been built on marginal cost estimates obtained from empirical models that impose strong parametric assumptions on cost functions. This paper evaluates these methods and, most importantly, proposes a better alternative. We first use actual data for marginal cost

and for variables required to estimate the marginal cost with a cost function. These data are from the well-established empirical industrial-organization literature that analyzes the electricity market, which is a unique case for observing marginal cost directly. Using this information, we also generate a panel data set from a suitable simulation process. Subsequently, we use both data sets to estimate the cost equation with (i) conventional parametric methods and (ii) three semiparametric and nonparametric models, namely the smooth coefficient model of Fan (1992), the generalized additive model of Hastie, and Tibshirani (1986) and the nonparametric inefficiency model of Kumbhakar et al. (2007).

We first show that conventional methods provide inaccurate estimates of marginal cost when applied to both the actual and the simulated data sets. This bias can be large and invalidate economic inferences. Next, we estimate the cost equation using the three semiparametric and nonparametric models. These models have a number of appealing features, the most important being that they allow to obtain coefficient estimates of marginal cost equal to the number of observations without imposing a specific functional form on the cost equation. We show that the smooth coefficient model and, most importantly, the generalized additive and the inefficiency models produce estimates of marginal cost that closely approximate the true marginal cost. In particular, (i) the correlation coefficient between the true and the estimated marginal costs is equal to or higher than 95% in all estimated alternative specifications and (ii) the probability density function of the estimated marginal cost maps very closely to that of the true marginal cost. We contend that these findings call for a reconsideration of the literature that relies on marginal cost estimates with a basis of parametric assumptions. This is of special relevance to the literature involving identification of industry conduct.

Further work needs to be done on the estimation of cost equations in other cases. For example, it would be interesting to re-evaluate the marginal cost measures within New-Keynesian macroeconomic models. This is another source of debate and friction within the economics literature, because of the requirement to estimate marginal cost for the aggregate economy. Moreover, this paper's framework can be used to re-examine the elements of profit maximization by equating the estimates of marginal cost and marginal revenue. Finally, the present analysis needs to be augmented by proposing a consistent estimator of marginal cost when the firm's output is endogenous. This augmentation calls

for an estimation method that uses local instrumental or local generalized method of moments techniques. We leave these as a desideratum for future research.

Acknowledgments

We are grateful to Severin Borenstein, James Bushnell, and Frank Wolak for providing their data from the California's electricity market. We are also grateful to two anonymous referees and Chung-Piaw Teo (Area Editor) for invaluable comments and suggestions. We also thank Iftekhar Hasan, Sofronis Clerides and Eugenia Vella, as well as seminar participants at Cass Business School, Brunel University, Surrey Business School and Athens University of Economics and Business. Delis acknowledges financial support from City University and University of Surrey and Tsionas from the Athens University of Economics and Business. The usual disclaimer applies.

References

- Aigner, D. J., C. A. K. Lovell, P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6, 21-37.
- Borenstein, S., J. B. Bushnell, F. A. Wolak. 2002. Measuring market inefficiencies in California's restructured wholesale electricity market. *American Economic Review* 92, 1376-1405.
- Bresnahan, T.F. 1989. Empirical studies of industries with market power. R. Schmalensee, R. D. Willig, eds. *Handbook of Industrial Organization*, vol. 2, Elsevier, Amsterdam, 1011-1057.
- Fan, J. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998-1004.
- Genesove, D., W. P. Mullin. 1998. Testing static oligopoly models: Conduct and cost in the sugar industry, 1890-1914. *Rand Journal of Economics* 29, 355-377.
- Hall, R. E. 1988. The relation between price and marginal cost in U.S. industry. *Journal of Political Economy* 96, 921-947.
- Hastie, T. J., R. J., Tibshirani. 1986. *Generalized Additive Models*. *Statistical Science* 1, 297-318.
- Hoel, P.G., 1966. Introduction to mathematical statistics. John Wiley and Sons, New York.
- Hoover, D. R., J. A., Rice, C. O., Wu, L., Yang. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85, 809-822.
- Kim, D., C. R. Knittel. 2006. Biases in static oligopoly models? Evidence from the California electricity market. *Journal of Industrial Economics* 54, 451-470.
- Koetter, M., J. W. Kolari, L. Spierdijk. 2012. Enjoying the quite life under deregulation? Evidence from unadjusted Lerner indices for U.S. banking. *Review of Economics and Statistics* 94, 462-480.
- Konings, J., P. Van Cayseele, F. Warzynski. 2005. The effects of privatizations and competitive pressure on firms' price-cost margins: Evidence from emerging economies. *Review of Economics and Statistics* 87, 124-134.
- Kumbhakar, S. C., B. U. Park, L. Simar, E. G. Tsionas. 2007. Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics* 137, 1-27.

- Kumbhakar, S., E. G. Tsionas. 2008. Scale and efficiency measurement using a semiparametric stochastic frontier model: Evidence from the U.S. commercial banks. *Empirical Economics* 34, 585-602.
- Li, Q., C. J. Huang, D. Li, T Fu. 2002. Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics* 20, 412-422.
- Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69, 307-342.
- Ramsey, J. B. 1969. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series B* 31, 350-371.
- Reiss, P. C., F. A. Wolak. 2007. Structural econometric modeling: Rationales and examples from industrial organization. J. J. Heckman, E. E. Leamer, eds. *Handbook of Econometrics*, vol. 2, Elsevier, Amsterdam, 4277-4415.
- Roeger, W. 1995. Can imperfect competition explain the difference between primal and dual productivity measures? Estimates for U.S. manufacturing. *Journal of Political Economy* 103, 316-330.
- Rosse, J. N. 1967. Daily newspapers, monopolistic competition, and economies of scale. *American Economic Review* 57, 522-533.
- Wolfram, C. D. 1999. Measuring duopoly power in the British electricity spot market. *American Economic Review* 89, 805-826.
- Wood, S.N. 2006. Generalized additive models: An introduction with R. Chapman and Hall, London.
- Zhang, W., S. Lee, X. Song. 2002. Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis* 82, 166-188.

Table 1. Summary statistics of actual data

Variable	Mean	Std. dev.	Min.	Max.
mc	34.45	17.97	18.68	441.11
png	27.30	9.02	15.65	47.42
pno	3.87	8.16	0.13	36.59
q	5,006.8	3,215.3	539.0	15,632.5
tc	181,056.7	179,676.7	12,914.9	1,067,486.0

Notes: The table reports summary statistics for the actual data. mc is the true marginal cost, png is the average daily price of natural gas for California, pno is the price of NO_x permits, q is the quantity of output produced and tc is the total cost. The data are hourly time series (number of observations is 21,217).

Table 2. Estimation results from the parametric models using actual data

	(1)	(2)	(3)	(4)	(5)
Functional form:	log-linear	translog	log-linear	translog	squared terms
Estimation method:	OLS	OLS	2SLS	2SLS	OLS
ln(png/pno)	1.016 (926.31)	0.896 (51.11)	1.009 (892.36)	-0.111 (-2.57)	0.770 (221.09)
lnq	1.105 (547.17)	0.538 (13.55)	1.069 (448.99)	-0.797 (-10.87)	0.435 (14.17)
$\frac{1}{2}[\ln(png/pno)]^2$		0.084 (52.19)		0.130 (52.21)	
$\frac{1}{2}(\ln q)^2$		0.071 (15.37)		0.188 (24.34)	
ln(png/pno)*lnq		-0.013 (-7.39)		0.089 (20.58)	
$[\ln(png/pno)]^2$					0.045 (71.43)
$(\ln q)^2$					0.039 (21.02)
Constant	-0.788 (-42.82)	1.669 (9.61)	-0.459 (-21.30)	8.996 (25.17)	2.306 (18.41)
Observations	21,217	21,217	21,217	21,217	21,217
Adj. R-squared	0.977	0.981	0.977	0.978	0.981

Notes: The table reports the results (coefficients and *t*-statistics) from the estimation of Equations (2) and (3) using the actual data from Borenstein et al. (2002). All coefficient estimates are statistically significant at the 1% level. The dependent variable is the natural logarithm of total cost divided by the price of NO_x permits, i.e. ln(tc/pno). png is the average daily price of natural gas for California, and q is the quantity of output produced at each point in time.

Table 3. Estimates of marginal cost from parametric models using actual data

Panel A. Summary statistics of estimates of marginal cost				
	Mean	Std. dev.	Min.	Max.
mc _{llols}	34.71	12.78	20.88	81.17
mc _{trols}	34.68	14.60	18.25	88.59
mc _{ll2sls}	35.27	12.99	21.22	82.48
mc _{tr2sls}	32.82	11.70	14.14	80.18
mc _{stols}	34.45	22.36	19.64	88.74

Panel B. Correlations between true marginal cost and estimated marginal costs from the different parametric models					
	mc _{llols}	mc _{trols}	mc _{ll2sls}	mc _{tr2sls}	mc _{stols}
mc	0.898	0.902	0.898	0.879	0.912

Notes: Panel A reports summary statistics for the estimates of marginal cost obtained from the four alternative parametric models of Table 2. Panel B reports pairwise correlation coefficients between the true values of marginal cost (mc) and the estimates of marginal cost obtained from the five alternative specifications in Table 2. mc_{llols} is from the estimation of the log-linear specification with OLS, mc_{trols} is from the estimation of the translog specification with OLS, mc_{ll2sls} is from the estimation of the log-linear specification with a two-stage least squares, mc_{tr2sls} is from the estimation of a translog specification with a two-stage least squares, and mc_{stols} is from the estimation with OLS of a specification with squared terms on the explanatory variables.

Table 4. Estimation results from the semiparametric and the nonparametric models using actual data

Panel A. Smooth coefficient model				
	Mean	Std. dev.	95% conf. interval	
png	2,689.9	441.7	2,575.7	2,804.3
pno	5,607.4	543.3	5,481.4	5,733.3
q (mc _{scm})	35.10	5.86	34.88	35.31
Correlation coefficient between mc _{scm} and mc = 0.974.				
Panel B. Generalized additive model				
	Mean	Std. dev.	95% conf. interval	
png	2,687.3	440.5	2,573.1	2,801.7
pno	5,606.2	537.8	5,480.1	5,732.4
q (mc _{gam})	34.71	5.71	34.49	34.94
Correlation coefficient between mc _{gam} and mc = 0.983.				
Panel C. Inefficiency model				
	Mean	Std. dev.	95% conf. interval	
png	2,669.7	443.4	2,551.1	2,784.2
pno	5,594.2	548.2	5,468.4	5,720.1
q (mc _{im})	34.63	5.76	34.41	34.86
Correlation coefficient between mc _{im} and mc = 0.987.				

Notes: The table reports means, standard deviations and bootstrap confidence intervals for the estimates from the smooth coefficient model (Panel A), the generalized additive model (Panel B) and the inefficiency model (Panel C). Here we use the actual data from Borenstein et al. (2002). png is the average daily price of natural gas for California, pno is the price of Nox permits, and q is the quantity of output produced at each point in time that also equals the estimated marginal cost.

Table 5. Summary statistics of simulated data

Variable	Mean	Std. dev.	Min.	Max.
mc	34.40	4.90	20.28	61.23
png	27.24	7.40	6.42	65.17
pno	3.85	2.78	0.01	32.93
q	5,007.1	99.8	4,599.5	5,390.3
tc	218,136.7	36,870.9	143,711.3	328,261.3

Notes: The table reports summary statistics for the simulated panel data. mc is the true marginal cost, png is the average daily price of natural gas for California, pno is the price of NO_x permits, q is the quantity of output produced and tc is the total cost. The panel data include information for 4,000 time series over 5 firms (20,000 observations).

Table 6. Estimation of parametric models using simulated data

	(1)	(2)
Functional form:	log-linear	translog
Estimation method:	OLS	OLS
$\ln(\text{png}/\text{pno})$	0.898*** (371.11)	-2.133** (-2.07)
$\ln q$	0.941*** (8.99)	41.048 (0.65)
$\frac{1}{2}[\ln(\text{png}/\text{pno})]^2$		0.074*** (21.60)
$\frac{1}{2}(\ln q)^2$		-4.795 (-0.65)
$\ln(\text{png}/\text{pno}) \cdot \ln q$		0.334*** (2.77)
Constant	1.219 (1.37)	-166.26 (-0.62)
Observations	20,000	20,000
Adj. R-squared	0.862	0.865

Notes: The table reports the results (coefficients and *t*-statistics) from the estimation of Equations (2) and (3) using the simulated data and OLS. The dependent variable is the natural logarithm of total cost divided by the price of Nox permits, i.e. $\ln(\text{tc}/\text{pno})$. png is the average daily price of natural gas for California, and q is the quantity of output produced at each point in time. The *** and ** denote statistical significance at the 1 and 5% levels respectively.

Table 7. Estimates of marginal cost from parametric models using simulated data

Panel A. Summary statistics of estimates of marginal cost				
	Mean	Std. dev.	Min.	Max.
mc _{llols}	41.70	7.02	28.53	62.33
mc _{trls}	42.37	14.12	6.64	139.81

Panel B. Correlations between true marginal cost and estimated marginal costs from the different parametric models		
	mc _{llols}	mc _{trls}
mc	-0.005	0.342

Notes: Panel A reports summary statistics for the estimates of marginal cost obtained from the two alternative parametric models of Table 5. Panel B reports pairwise correlation coefficients between the true values of marginal cost (mc) and the estimates of marginal cost obtained from the two alternative specifications in Table 5. The mc_{llols} is from the estimation of the log-linear specification with OLS, and mc_{trls} is from the estimation of the translog specification with OLS.

Table 8. Estimation results from the semiparametric and the nonparametric models using simulated data

Panel A. Smooth coefficient model				
	Mean	Std. dev.	95% conf. interval	
png	2,645.3	310.5	2,528.1	2,762.7
pno	5,355.1	402.2	5,225.0	5,485.3
q (mc _{scm})	34.70	2.47	34.48	34.94
Correlation coefficient between mc _{scm} and mc = 0.950.				
Panel B. Generalized additive model				
	Mean	Std. dev.	95% conf. interval	
png	2,647.1	308.44	2,529.9	2,764.5
pno	5,349.2	406.52	5,219.2	5,479.4
q (mc _{gam})	34.48	2.40	34.26	34.71
Correlation coefficient between mc _{gam} and mc = 0.966.				
Panel C. Inefficiency model				
	Mean	Std. dev.	95% conf. interval	
png	2,658.5	314.31	2,541.3	2,775.6
pno	5,388.2	381.20	5,271.0	5,505.5
q (mc _{im})	34.54	2.30	34.32	34.76
Correlation coefficient between mc _{im} and mc = 0.944.				

Notes: The table reports means, standard deviations and bootstrap confidence intervals for the estimates from the smooth coefficient model (Panel A), the generalized additive model (Panel B) and the inefficiency model (Panel C). Here we use the simulated data, constructed on the basis of the actual the data set of Borenstein et al. (2002). png is the average daily price of natural gas, pno is the price of Nox permits, and q is the quantity of output produced at each point in time that also equals the estimated marginal cost.

Figure 1. Total cost and electricity output

Figure 2. Total cost and the price of natural gas

Figure 3. Total cost and the price of NOx

Figure 4. Marginal cost obtained from the log-linear specification and OLS vs. true marginal cost (actual data)

Figure 5. Marginal cost obtained from the translog specification and OLS vs. true marginal cost (actual data)

t-test of equality: 4.14 ($p=0.000$), sign-test: $p=0.000$

Figure 6. Marginal cost obtained from the a log-linear specification and 2SLS vs. true marginal cost (actual data)

t-test of equality: 14.02 ($p=0.000$), sign-test: $p=0.000$

Figure 7. Marginal cost obtained from the translog specification and 2SLS vs. true marginal cost (actual data)

t-test of equality: 25.06 ($p=0.000$), sign-test: $p=0.000$

Figure 8. Marginal cost obtained from the model with squared terms vs. true marginal cost (actual data)

t-test of equality: 2.71 ($p=0.001$), sign-test: $p=0.000$

Figure 9. Marginal cost obtained from the smooth coefficient model vs. true marginal cost (actual data)

Figure 10. Marginal cost obtained from the generalized additive model vs. true marginal cost (actual data)

Figure 11. Marginal cost obtained from the nonparametric inefficiency model vs. true marginal cost (actual data)

Figure 12. Marginal cost obtained from the log-linear specification and OLS vs. true marginal cost (simulated data)

Figure 13. Marginal cost obtained from the translog specification and OLS vs. true marginal cost (simulated data)

Figure 14. Marginal cost obtained from the smooth coefficient model vs. true marginal cost (simulated data)

Figure 15. Marginal cost obtained from the generalized additive model vs. true marginal cost (simulated data)

Figure 16. Marginal cost obtained from the nonparametric inefficiency model vs. true marginal cost (simulated data)