Panel Data in Energy Economics*

Appearing in: Panel Data Econometrics Volume 2: Empirical Applications, ed. Mike Tsionas,
Elsevier Publishing Co.

Scott E. Atkinson
Department of Economics
University of Georgia
email: atknsn@uga.edu

January 4, 2018

—I wish to thank Chris Cornwell, Rong Luo, Tom Quan, and Mike Tsionas, for helpful comments on an earlier draft.

†Corresponding author with phone: 706 549 0977; fax: 706 542 3376
Abstract

This chapter analyzes panel data studies of the most widely-examined energy consumption industries—electric power, railroads, and airlines. For electric power, the choice between utility-level versus plant-level data, cross-sectional versus panel data, and pooled-data analysis versus fixed-effects (FE) estimation generally makes little difference. A consensus also exists across estimates of cost, profit, and distance functions as well as systems including these functions. Generally, studies reject homogeneous functional forms and find nearly constant returns to scale (RTS) for the largest firms. Residual productivity growth declines over time to small, positive levels and substantial economies of vertical integration exist. Cost saving may accrue from a competitive generating sector. Controversy remains regarding the Averch-Johnson effect and the relative efficiency of publicly-owned versus privately-owned utilities. Railroads exhibit increasing RTS, substantial inefficiencies, and low productivity growth. Airlines operate close to constant RTS and enjoy modest productivity growth. Substantial inefficiencies decrease with deregulation. A valuable alternative to FE estimation is a control function approach to model unobserved productivity.

KEY WORDS: panel data, firm fixed effects, railroads, airlines, electric utilities, production functions, cost functions, profit functions, shadow costs, distance functions
1 Introduction

In 2016, the electric-power sector consumed 39% of all energy resources, the transportation sector 29%, the industrial sector 22%, and the residential/commercial sector 11%. The use of panel data to study energy economics is so vast that this chapter focuses on econometric analyses since the 1960s of three of the largest and most widely-studied consumers of energy resources—the fossil-fueled electric-power sector, railroads, and airlines.

To fully understanding the advantages (and potential pitfalls) of panel data analysis, we review the cross-sectional estimation of the production, input demand, cost, and profit functions for electric utilities that began in the 1960s. Analysts rightly criticize early production function estimation for using homogeneous functional forms, ignoring the endogeneity of explanatory variables, and often employing macro data which obscures causality. Responding to these criticisms, researchers in the 1970s use cross-sectional micro data to estimate flexible functional forms for input demand, cost, and profit systems (where cost and profit systems include derived demand or share equations). These analysts assume that arguments of their functions (inputs prices, output prices, and output quantities) are exogenous. Panel data in the late 1970s allows the calculation of the fixed effects (FE) estimator, which eliminates time-invariant unobserved heterogeneity and hence reduces the potential for endogeneity. Panel data also enables the calculation of firm- and input-specific price-inefficiency parameters. Further, the researcher can compute firm-specific measures of technical efficiency (TE), technical change (TC), efficiency change (EC), and productivity change (PC). While TE measures the distance of the firm from the production frontier, EC is the change in this measure over time, TC is the shift in the frontier over time, and PC is the sum of the latter two measures. In the 1980s interest focuses on multi-product cost functions and economies from vertical integration (EVI) of production, transmission, and generation of electricity.

In the late 1980s econometricians return in full force to estimate production functions. Now researchers use panel data and specify multiple-output production functions. Some formulate distance and directional distance functions for electric utilities, where multiple outputs are either good (such as residential and commercial/industrial generation) or bad (such as SO$_2$, CO$_2$, and NO$_x$ emissions). Researchers often avoid direct estimation of cost and profit functions, since many good inputs (such as capital, labor, and energy) and bad outputs lack market prices. When exogenous prices are available, some analysts specify distance and directional distance systems (distance and directional distance functions together with first-order conditions from cost-minimization or profit-maximization problems). Typically, researchers compute an instrumental variable (IV) estimator, where prices comprise a subset
of their instruments. In some cases, econometricians employ Bayesian analysis via Gibbs sampling with instruments to facilitate the imposition of constraints, estimation of input- and firm-specific price inefficiencies, and residual-based calculation of firm-specific TE, EC, TC, and PC. Many researchers formulate multiple-output production functions for railroads and airlines. Other studies introduce control functions into production functions to directly model unobserved productivity, as part of a potential solution to the endogeneity problem. This approach is especially useful to achieve identification in the absence of exogenous input and output prices. Further, this approach improves upon the older residual-based method for calculating PC and its associated measures. However, since the control-function approach requires strong assumptions to achieve identification, it is another tool rather than a panacea in the treatment of endogeneity. More recently, a number of studies use macro data, employ homogeneous functional forms, or ignore the potential endogeneity of inputs. In these cases, we have come full circle to repeat the errors of the production function estimation from the 1960s, now using panel rather than cross-sectional data.

Substantial agreement exists among studies of the electric power, railroad, and airline industries. For the electric power sector, this is true regardless of the choice between utility-level versus plant-level data, cross-sectional versus panel data, and pooled-data analysis versus FE estimation. Further, a general consensus exists across estimates of cost, profit, and distance functions, as well as systems including these functions. Scale economies appear to be exhausted for the largest firms, rates of productivity growth are moderate, and substantial EVI exists for the generation, transmission, and distribution sectors. Further, cost saving may accrue from a competitive generating sector. Although allocative inefficiency appears to exist, substantial disagreement persists regarding the existence and magnitude of an Averch and Johnson (1962) (AJ) effect and the relative efficiency of public versus private utilities. The AJ effect occurs when the firm over-capitalizes due to rate-of-return regulation, so long as the allowed rate of return exceeds the market return on capital. Railroads exhibit increasing returns to scale (RTS), substantial inefficiencies, and low productivity growth. Airlines operate close to constant RTS and enjoy modest productivity growth. Substantial inefficiencies decrease with deregulation. A valuable alternative to FE is a control function approach to model productivity, which allows one to control for endogeneity and compute partial effects from a productivity function.
2 Cross-Sectional Modeling of Electric Utilities

2.1 Production Functions without Instruments

Initial examination of a number of cross-sectional studies allows comparison with the results from later panel studies of similar topics. The general consensus from these cross-sectional studies is that RTS are nearly exhausted for the largest firms, that allocative inefficiency is moderate for some inputs, that the AJ effect exists, and that EVI are substantial. Douglas (1976) references early estimates of the Cobb-Douglas production function using macro data without concern for endogeneity of inputs, a practice that is criticized by Griliches and Mairesse (1998) (GM). Cowing and Smith (1978) survey a number of early studies by Komiya (1962), Galatin (1968), and Courville (1974) that utilize micro cross-sectional data on US utilities for various time periods to fit electric utility production functions. However, they estimate a homogeneous production function and ignore the endogeneity of inputs. Generally they obtain evidence of substantial technical progress and moderate scale economies.

2.2 Cost, Derived Demand, and Profit Functions

Nerlove (1963) is among the first to address endogeneity by recognizing that the inputs of an electric utility production function are endogenous. Instead, he estimates a Cobb-Douglas cost function, whose arguments are output quantity and input prices for capital, labor, and energy. These variables are arguably exogenous for a regulated firm buying inputs in competitive markets and selling output at prices determined by regulators. Using a cross-sectional sample of 145 US firms in 1955, Nerlove obtains evidence of a marked degree of increasing RTS for smaller firms, where \( \text{RTS} = 1 - (\partial \ln C / \partial \ln y) \), \( C \) is cost, and \( y \) is output. Thus, positive numbers indicate scale economies and negative numbers indicate scale diseconomies. He finds that RTS falls from .91 to -.03 as output increases. Extending this work, Dhrymes and Kurz (1964), Cowing (1974), and Boyes (1976) use cross-sectional data on new US plants installed over time to estimate more flexible input demand models, assuming either cost-minimization or profit-maximization. The consensus result is that RTS declines with firm size to nearly zero.

One of the first studies to employ a second-order flexible functional form is Christensen and Greene (1976), who specify a translog cost system for cross sections of US utilities in 1955 and 1970. They find significant increasing RTS for nearly all firms in the earlier group, but essentially constant RTS for the later group of larger firms. They reject the null hypotheses of homothetic and homogeneous technologies. Nearly all studies from this point forward specify a second-order flexible functional form.
for cost, profit, distance, and directional distance functions. Only the use of the Cobb-Douglas function will warrant specific mention. In a related study, Christensen and Greene (1978) estimate a cost system for 138 US firms in 1970 and find no evidence of systematic cost savings from power pooling. However, commonly-owned firms exhibit lower average costs than comparable individual firms.

Atkinson and Halvorsen (1984) generalize this work by formulating a shadow cost system for 123 US private utilities in 1970, where shadow costs are a parametric function of shadow prices. These are virtual prices to the firm that may differ from actual or market prices due to regulatory constraints, labor union rules, or other forces that impede free markets from efficiently setting prices. Using cross-sectional data, shadow cost systems (and profit systems to be addressed shortly) allow direct estimation of input-specific allocative efficiency (AE) parameters. Atkinson and Halvorsen find that RTS are nearly exhausted for the largest firms, evidence of an AJ effect, and moderate allocative inefficiency. Two studies formulate shadow cost systems to examine the relative efficiency of public and private utilities using cross-sectional data. Atkinson and Halvorsen (1986) employ 1970 data for 123 private and 30 public US electric utilities, concluding that the two ownership types are equally cost efficient. Other cross-sectional studies summarized in their paper are divided in their assessments of relative efficiency. Following the methodology of Atkinson and Halvorsen (1984), Koh, Berg, and Kinney (1996) utilize a 1986 cross-section of 121 private and 61 public US utilities. They find that public utilities are more efficient at low output levels.

A number of studies estimate profit functions as a dual alternative to cost functions. Atkinson and Halvorsen (1976) specify a normalized restricted profit system (a profit function plus derived share equations), whose arguments are input prices, output prices, and quantities of fixed inputs. They assumed exogeneity of these prices, since arguably individual plants are price takers in both input and output markets. For 108 US power plants in 1972, they find that RTS are very close to zero. Cowing (1978) estimates a profit system for 150 new private US plants constructed between 1947 and 1965. He provides general support for the effectiveness of rate-of-return regulation and the existence of an AJ effect. Atkinson and Halvorsen (1980) generalize previous work by estimating AE parameters using a normalized shadow profit system for a set of 38 US power plants in 1973. Shadow profits are a parametric function of shadow prices. They find that these plants do not generally achieve relative and absolute price efficiency.

Meyer (2012a) surveys a number of cross-sectional studies of EVI for US utilities. The consensus conclusion is that separating generation from transmission, distribution, and retail services substantially increases costs. Using data on US firms, Roberts (1986), Kaserman and Mayo (1991), and Gils-
hof (1995) generally agree that a reduction in average cost results from increasing output to existing customers, with no substantial savings from increasing the number of customers. Further, they find substantial EVI. Huettner and Landon (1978) estimate cost functions for production, transmission, and distribution for a cross section of 74 US electric utilities in 1971. They find diseconomies of scale beyond moderate firm sizes for all three activities. They also find that holding companies do not generate cost savings. Using data for 1999, Bushnell, Mansur, and Saravia (2008) explain the cost, price, and residual supply of electricity for three market areas in the US. They find that the prevention by regulators of the long-term price commitments associated with vertical arrangements (as in California) increases prices by 45%.

3 Pooled and Fixed-Effects Panel Data Modeling

Most research on electric utilities employs panel data. Consider the estimation of a linear panel data model for firm $i$ in period $t$:

$$y_{it} = x_{it}\beta + c_i + e_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T;$$

(1)

where $y_{it}$ is the dependent variable, $x_{it}$ is a $(1 \times K)$ vector of regressors, $\beta$ is a $(K \times 1)$ coefficient vector, and the error term is comprised of $c_i$, which is firm-specific, time-invariant unobserved heterogeneity and an idiosyncratic error term, $e_{it}$.

One advantage of panel over cross-sectional data occurs in the treatment of endogeneity. Endogeneity occurs if $c_i$ or $e_{it}$ is correlated with the explanatory variables. With panel data one can remove $c_i$ from the error term by computing the FE estimator of $\beta$, through time-demeaning the data or estimating $c_i$ directly. With a linear model, the FE estimator is also termed the “within” estimator. Alternatively one can assume away endogeneity by employing the random-effects (RE) estimator, which requires that the composite error term, $c_i + e_{it}$, is orthogonal to $x_{it}$. Using linear equations analogous to (1), the analyst can compute the within estimator for linear cost, profit, or distance functions. For cost, profit, or distance systems, one directly estimates $c_i$ using FE. Clearly, FE methods do not eliminate the time-varying correlation between $e_{it}$ and $x_{it}$. As an alternative to FE and RE estimation, one can simply pool the untransformed panel data (ignoring $c_i$) and compute what is termed the “pooled” estimator.

A second advantage of panel over cross-sectional data is in the calculation of firm-level productivity growth and efficiency over time. Early cross-section estimates of cost and production functions do compare estimated models for each cross section of firms over time. However, panel data allows efficiency...
gains in estimation and the computation of firm-specific parameters determining AE, EC, TC, and PC. Despite the ability of the FE estimator to remove $c_i$ from the error term, GM argue against its use. They provide evidence that FE estimates of RTS vary substantially from those obtained using pooled panel data. They argue that the FE estimator does more harm than good by eliminating a rich source of variation, failing to eliminate time-varying endogeneity, and increasing measurement error. Also, they speculate that the econometrician may more precisely estimate productivity by avoiding the FE estimator in favor of a control function approach, which we consider later.

3.1 Electric Utilities

3.1.1 Single-Output Input Demand, Cost, and Production Functions

Early studies using panel data either specify input demand equations, presumably avoiding endogeneity, or estimate production functions while ignoring endogeneity. Barzel (1964) employs a panel of 220 US plants from 1941-59 and uses pooled data to compute input demand equations as functions of price and other controls. He obtains evidence of increasing RTS which are most pronounced for labor and less so for capital and fuel. On this basis, he argues that a Cobb-Douglas production function is inadequate for study of the electric power industry. Belinfante (1978) estimates a production function using a pooled unbalanced panel of 80 US electric utility plants from 1947-59 and calculates that average embodied TC is 3% and disembodied TC is .5%. This study ignores endogeneity.

Many panel data studies of single-output production by private electric utilities since the 1980s formulate pooled cost system models. They universally assume that the arguments of the cost function are exogenous. A nearly universal consensus from panel and previously summarized cross-sectional studies is that RTS are close to zero for the largest firms. In addition, the panel data studies generally find that productivity growth is slowing and that TC has become small over time. Petersen (1975) estimates a cost function for a pooled panel of 56 US private plants that experience substantial expansion from 1966-68. He finds evidence of an AJ effect and increasing RTS for all but the largest plants. Gollop and Roberts (1981) formulate a cost system for a pooled sample of 11 private, vertically-integrated US firms from 1958-75. They calculate that increasing RTS range from .31 to .1, but that RTS falls substantially in the last year. The rate of TC generally declines over their sample to negative levels in this year. Nelson and Wohar (1983) estimate total factor productivity (TFP) growth using a cost system by pooling data on 50 private US utilities from 1950-78, where TFP is the ratio of real output to a Divisia index of real factor inputs. Including a measure of the allowed rate of return, they find
evidence of an AJ effect, consistent with cross-sectional studies references above. Additionally they estimate that TC is the primary source of TFP growth which averaged 2.5%, and that RTS is stable over time at small positive levels. In contrast, Baltagi and Griffin (1988) compute a general index of TC after estimating a cost system using panel data with firm FE for 30 private US utilities from 1951-78. They first calculate TC (which is the negative of the difference in log costs between the present and a previous period) from 1971-78. Then they regress this measure on vintage, SO2 restrictions, regulatory tightness, and capacity utilization. The biggest factor decreasing TC is a restriction on SO2 production, enacted in the 1970’s, along with reduce utilization rates. However, since these measures are arguably correlated with variables in the first-stage cost function, the caveats of Wang and Schmidt (2002) (WS) apply. They show via Monte Carlo simulation that regression of first-step residuals on second-step explanatory variables—when these are correlated with first-step explanatory variables but omitted from that step—generates substantial bias in both steps. Using a cost system and a pooled panel from 1995 and 1996 for 775 US private plants, Maloney (2001) estimates that the largest plants exhaust RTS. Law (2014) summarizes a number of studies of US utilities that examine the existence of an AJ effect and criticizes most studies that confirm its existence.

Three studies examine the efficiency of public utilities or compare the efficiencies of public and private utilities. Nelson and Primeaux (1988) examine a pooled panel of 23 US municipally-owned utilities from 1961-76. They formulate a cost function for total distribution costs and find that substantial scale economies with respect to output exist throughout most of the sample. However, they calculate that RTS with respect to simultaneous increases in output and the number of customers appear to have been exhausted by the larger firms. This result is consistent with cross-sectional studies of US utilities by Huettner and Landon (1978) and by Roberts (1986). Pescatrice and Trapani (1980) estimate a cost system for a pooled sample of 23 public and 33 private US utilities from 1965-76. In contrast to Atkinson and Halvorsen (1986), they calculate that public firms minimize costs and have 24-33% lower per unit costs than private firms. They attribute the cost differential to rate-of-return regulation of the latter. Nemoto, Nakanishi, and Madono (1993) formulate a variable cost system using pooled data for 9 vertically-integrated Japanese firms from 1981-83. They obtain evidence of short-run economies of scale and long-run diseconomies of scale plus evidence supporting the AJ effect.

Restructuring the generation sector so that it is competitive appears to reduce costs. Rather than estimate a production function directly, Fowlie (2010) examines the effect of restructuring on plant input efficiency and emissions of NOX, using a pooled panel of 702 coal-fired US electric generating plants from 2000-04. Estimates from a random coefficients logit model indicate that deregulated plants
in restructured electricity markets are less likely to install more capital intensive pollution control technologies compared to similar regulated and public plants. Further, damages to health and the environment would have been less under symmetric economic regulation (universal restructuring or universal regulation), with relatively more of the permitted NO\textsubscript{x} emitted in relatively low damage areas. In a review of papers examining the effects of electricity restructuring on rates, Hiebert (2002) specifies a variable cost stochastic frontier (with a two-component error term as in (1)) for an unbalanced panel of 432 plants burning coal and 201 plants burning natural gas and oil in the US from 1988-97. He rejects the Cobb-Douglas form and finds evidence of greater efficiency as plant utilization rises, as the number of plants owned by a utility increases, as investor-owned status increases, and as coal plants are restructured. Kwoka (2008) critiques a number of panel data studies of restructuring. One of the more convincing studies is by Joskow (2006), who examines nearly all US states for 34 years. Estimating price regressions which include state FE, he finds that restructuring significantly reduces the price of electricity in the residential and industrial markets by 5% to 10%. Rungsuriyawiboon and Stefanou (2007) formulate a dynamic FE efficiency model of shadow cost minimization for 72 private US utilities from 1986-99. They treat net investment demand as endogenous and find that TE of inputs improves for utilities in restructured jurisdictions. They also determine that capital is overutilized relative to other inputs, but differences in allocative inefficiencies of variable inputs are not significantly different for deregulated versus regulated utilities. Estimated RTS (defined as the elasticity of output with respect to all inputs) ranges from 1.22 to 1.37, with higher RTS for the pre-deregulation period.

Two subsequent studies employ an IV estimator with FE for distance systems. A popular choice to model single-output production is the input-oriented distance function, which is dual to a cost function. Atkinson and Dorfman (2009) specify an input distance system with firm FE to explain total generation for a panel of 13 Chilean hydroelectric power plants from 1986-97. Panel data allows calculation of a full set of input- and plant-specific AE parameters using Gibbs sampling, which draws sequentially from conditional Generalized Method of Moments (GMM) IV estimates. They compute substantially differing degrees of efficiency across plants, with measures of PC ranging from 2% to 12%. Over time TE improves for most plants. Using data for 78 US private utilities from 1988-97, Atkinson and Cornwell (2011) estimate firm FE using a dynamic shadow input distance system that integrates dynamic adjustment costs into a long-run shadow cost-minimization problem. They formulate a system comprised of the first-order conditions from the short-run shadow cost-minimization problem for the variable shadow inputs, a set of Euler equations, and the input distance function, expressed in terms of shadow quantities. Estimates indicate that adjustment costs represent about 0.42% of total cost.
and about 1.26% of capital costs. Instruments include exogenous input prices. Compared to the static model, the dynamic model finds that over-use of capital and labor is less severe and annual PC is more stable and higher, at about 1.7% on average.

Llorca, Orea, and Pollitt (2016) examine the effect of environmental factors (such as wind and precipitation) on the efficiency of the US electricity transmission industry for the period 2001-09, using a pooled panel of 59 US electricity transmission companies. The authors recognize the caveats of WS and examine a number of alternative random-effects frontier models where firm inefficiency is a function of external variables such as weather. They find that the cost elasticity with respect to network length evaluated at the sample mean is 0.89. More than half of the firms in their sample exhibit increasing RTS based on this measure. Further, the cost elasticity with respect to density ranges from .70 to .75. Together these imply the existence of important economies of density in electricity transmission. Finally, they find that adverse weather reduces firm efficiency.

A number of recent studies ignore the concerns of GM regarding the use of macro variables and endogeneity in production functions. Zhang, Parker, Kirkpatrick (2008) assess the effects of privatization, competition, and regulation on the performance of the electricity generation industry using panel data for 36 developing and transitional countries from 1985 to 2003. They estimate country FE for a model where the dependent variables are electricity generation and explanatory variables measure regulation, competition, privatization, and other controls. However, they do not address potential remaining endogeneity. Pompei (2013) regresses Malmquist measures of TFP on TC, EC, and measures of regulatory stringency in the electricity sectors of 19 EU countries from 1994-2007. He computes pooled regressions where many of the explanatory variables are arguably endogenous. Oh and Lee (2016) estimate a firm FE production function for a panel of 5 Korean utilities from 2001-12, without addressing potential remaining endogeneity. Growth of TFP is 0.33% and RTS are small but positive. Polemis and Stengos (2017) explain generation, capacity, and productivity using aggregate variables and country-specific FE for 30 OECD countries from 1975-2013. For already economically liberalized countries, the level of economic freedom does not affect levels of production. Again, the authors do not address potential remaining endogeneity and the use of macro data clouds causal inference.

3.1.2 Multiple—Output Cost and Production Functions

A large body of literature estimates two different types of multi-product cost systems. With the first system, the outputs are generation, transmission, and distribution (and sometimes retail) of electricity. With the second system, the outputs are residential, commercial, and industrial electricity generation,
possibly also including bad outputs. Most studies compute pooled cost system regressions, assuming that the arguments of their cost functions are exogenous. A few studies estimate production functions, employ instruments, and specify FE.

### 3.1.2.1 Cost Savings from Economies of Vertical Integration

A number of studies examine EVI for generation, transmission, and distribution of electricity in the US and other countries. The strong consensus is that substantial EVI exist. Ramos-Real (2005) provides a survey of cost function estimation, arguments in favor of restructuring, and evidence of EVI. In the first of a series of studies of EVI for the US, Thompson (1997) formulates a cost system based on the total cost of production, procurement, and delivery of power using a pooled panel consisting of all major investor-owned US electric utilities for the years 1977, 1982, 1987, and 1992. He calculates substantial scale economies for expanding sales (with a given customer base) but no economies for expanding the customer base (for a given level of sales). Further, he finds substantial EVI. Hayashi, Goo, and Chamberlain (1997) examine 50 private vertically-integrated US utilities from 1983-87 by estimating firm FE with a cost function for both electricity supply and generation. Estimates of scale economies for generation are about 10% for smaller firms, and fall by half for larger firms. They also reject the null hypothesis of separability of generation and transmission/distribution implying that vertical integration could reduce total costs. Meyer (2012b) estimates a FE cost function where generation, transmission, and distribution are outputs, using a sample of 143 US utilities from 2001-08. He determines that separating generation from networks and retail is the most costly alternative with an average cost increase of 19 to 26%. If generation and transmission remain integrated but separate from distribution and retail, costs would increase by 8 to 10%. Finally, separating transmission from the remaining supply stages, would increase costs by approximately 4%.

A number of studies of EVI focus on OECD and European countries. Hattori and Tsutsui (2004) analyze the impact of regulatory reform on the level of the industrial price and the ratio of the industrial to residential price, using panel data for 19 OECD countries from 1987-99. They calculate country-specific FE, finding that expanded retail access lowers the industrial price and increases the price differential between industrial and household customers, while unbundling generation from transmission and the introduction of a wholesale spot market do not necessarily lower the price and may possibly increase it. Steiner (2001) estimates firm FE to explain the impact of industry structure and regulation on efficiency and prices, using panel data for 19 OECD countries over the period 1986-96. However, the calculated benefits from unbundling are problematic, since explanatory variables are macro (which may
not be the actual causal factors) and the model ignores the potential endogeneity of many explanatory variables.

Other studies examine Spanish, Japanese, Italian, and New Zealand electric utilities, generally finding substantial EVI. Martínez-Budría, Jara-Díaz, and Ramos-Real (2003), Jara-Díaz and Ramos-Real (2004), and Jara-Díaz and Ramos-Real (2011) formulate aggregate, multi-output (generation and distribution), and multistage-multioutput (coal, oil, hydraulic, nuclear, and distribution) cost functions for the Spanish electric sector using a panel of 12 Spanish utilities from 1985-96. This sample includes a period of regulatory reform. Aggregation bias for the multistage-multioutput function, due to combining such disparate production technologies, is a paramount concern. They compute firm FE and find EVI of about 9% as well as essentially constant RTS at the multistage level. Nemoto and Goto (2004) formulate a shadow cost system using a pooled panel data set on the transmission/distribution stages of nine Japanese electric utilities from 1981-88. The results indicate EVI for these two stages. Further, they calculate a negative externality effect of generation facilities on the cost of the transmission/distribution stage. This implies that vertical integration of generation and transmission/distribution facilities will generate EVI by internalizing these externalities.

Fraquelli, Piacenza, and Vannoni (2005) specify a multiproduct cost system using a pooled sample of 25 Italian electric utilities from 1994-2000. They include generation and distribution as separate outputs, finding overall increasing RTS for large utilities due to the presence of vertical economies. The latter counterbalance the effects of decreasing RTS in the generation phase. For the average firm the authors find weak but statistically significant EVI (3%) as well as multi-stage RTS of about 1.015, using the same formula as Atkinson and Primont (2002), where

$$\text{RTS} = \frac{1}{\Sigma_m (\partial \ln C/\partial y_m)_y / y_m}, \quad m = 1, \ldots, M,$$

for $M$ outputs and $C$ is cost. Numbers greater (less) than one indicate increasing (decreasing) RTS. Utilities which generate and distribute more than average amounts of electricity benefit from both EVI and increasing RTS, while the cost advantages increase up to 40% for large operators.

Studies in New Zealand generally indicate substantial EVI. Scully (1998) formulates a pooled cost system for New Zealand electrical supply authorities from 1982-94. He includes controls for engineering characteristics, whether the supply authority is municipal, and the degree of horizontal integration, finding that deregulation substantially reduces costs, with the real price of electricity falling 16.4% over the period. Scale economies decline to small levels as output expands. Negative TC through 1984 increases to about 1% in 1994. Nillesen and Pollitt (2011) estimate firm FE and RE models using a Cobb-Douglas cost function for 28 New Zealand utilities from 1995-2007. They determine that deregulation swaps one form of vertical integration (retail-distribution) for another form of vertical
integration (retail-generation). Further, ownership unbundling in New Zealand substantially reduces costs and increases quality of service, while overall competition falls and prices rise.

The potential exists for endogeneity in cost system studies of EVI. In a cost-system framework, estimating input demand equations by regressing inputs on input prices and output may suffer from endogeneity. The production function specified by GM, p. 172, is

\[ y = \alpha z + \beta x + u, \]  

where \( y \) is the log of output, \( z \) the log of capital (or a fixed input), and \( x \) is the log of labor (or a variable input). They then assume that product prices are equal for different producers and hence are normalized to unity. Equating the price of labor to the marginal product of \( x \), derived from (2), and taking logs yields the following marginal productivity equation:

\[ y = x + w + v - \ln(\beta), \]  

where \( w \) is the log of the wage and \( v \) is the error in the marginal product equation.

Solving (2) and (3) for their reduced-form equations, GM obtain (dropping constant terms):

\[ x = (1 - \beta)^{-1}(\alpha z - w + u - v) \]  

\[ y = (1 - \beta)^{-1}(\alpha z - \beta w + u - \beta v). \]

Thus \( x \) and \( y \) are correlated with both \( u \) and \( v \), so that input demand equations, whose arguments are input prices and output, suffer from endogeneity. As GM point out, if utilities operate within a regulatory environment the assumption that \( y \) is exogenous may be valid. However, in a restructured environment, where output is no longer determined by a regulatory commission, output is most likely endogenous to the firm.

### 3.1.2.2 Multiple Products–Multiple Good Outputs or Good with Bads

A number of studies examine the joint production of good and bad outputs. Gollop and Roberts (1983) measure the effect of restrictions on \( \text{SO}_2 \) emissions for productivity growth in the electric power industry from 1973-79. They formulate a cost system for 56 pooled US private firms, assuming the exogeneity of the arguments of the cost function—prices of labor, capital, low-sulfur fuel, high-sulfur
fuel, output, regulatory intensity, and time. In addition to rejecting the null of constant RTS, they find that the smallest firms enjoyed substantial scale economies, while the largest firms exhibited scale diseconomies. The mean RTS in 1979 is about .08. They also conclude that the biggest factor reducing TC is the restriction on SO₂ emissions enacted in the 1970’s. Färe, Grosskopf, Noh, and Weber (2005) estimate an output-oriented directional distance function to measure the TE of 209 US utilities in 1993 and 1997 observed at the boiler level. Performing separate cross-section regressions, they model two outputs, electricity and SO₂, as a function of standard inputs. While their approach produces reasonable shadow prices for SO₂, they ignore endogeneity. Atkinson and Dorfman (2005) compute firm FE for an input-oriented distance system. They model a good output (electricity generation) and a bad output (SO₂ production) using a panel of 43 private US electric utilities for 1980, 1985, 1990, and 1995. Using Bayesian GMM with instruments, Gibbs sampling allows easy imposition of monotonicity. They compute reasonable shadow values for SO₂ and find declining levels of PC and TC in addition to negative EC. A substantial portion of PC and TC is due to a reduction of the bad. Atkinson and Tsionas (2016) use Bayesian methods to estimate optimal firm-specific directions for a technology-oriented directional distance system using FE. They examine 77 US utilities from 1988-97, where production is a function of good inputs and outputs as well as bad inputs and outputs (SO₂, CO₂, and NOₓ). Estimated firm-specific directions for each input and output are quite different from those normally assumed. The computed firm-specific TE, TC, and PC using estimated optimal directions are substantially higher than those calculated using fixed directions.

Three papers estimate multiple outputs for the residential the industrial/commercial sectors. Estimates of RTS and inefficiencies are in general agreement with those from previous studies. Atkinson, Cornwell, and Honerkamp (2003) estimate an input distance function with firm FE, using a GMM estimator with instruments, examining 43 private US utilities from 1961-92. Outputs are residential and industrial/commercial generation. Estimates of RTS range from .8 to 1.13, with an average of 1.04, using the Atkinson and Primont (2002) formula. They also find small positive weighted-average PC computed from residuals. Atkinson and Primont (2002) generalize previous work by estimating multiple-output cost and production systems. The multiple outputs are residential and commercial/industrial production. They specify firm FE for shadow cost and shadow distance systems. Their data is a balanced panel of 43 private US electric utilities from 1961-97. The authors include instruments and find moderate allocative inefficiency, evidence of an AJ effect, very small productivity growth, and increasing RTS from about 1.13-1.15. Hayashi, Sevier, and Trapani (1985) estimate a rate-of-return cost system for residential, commercial, and industrial electricity. They employ a pooled panel of 32 private,
vertically-integrated US firms in 1965 and 1970, finding evidence of an AJ effect. However, their results are problematic, since they did not impose linear homogeneity in prices, which is a required accounting identity for a cost function, either in actual prices or in shadow prices.

### 3.2 Railroads

Although a number of panel data studies examine the efficiency of railroads and their RTS, the only generally robust conclusions are that RTS are modest and that substantial differences in firm efficiencies exist. Kumbhakar (1987) compares the estimates of a Cobb-Douglas production system to those of a Cobb-Douglas cost system to determine the TE and AE for a set of 13 Class-I US railroads from 1951-75. He assumes that both outputs—freight-ton miles and passenger miles—are exogenous and that the production system (cost share equations estimated jointly with the production function) is free of endogeneity. This is in contrast to the conclusions of GM that are drawn from (4) and (5). Due to considerable differences between the results of the cost and production systems, Kumbhakar advocates against the latter.

Atkinson, Färe, and Primont (2003) include firm FE to formulate a shadow input distance system for 12 US Class-I railroads from 1951-75, using a GMM IV procedure. Their shadow input distance equation is a function of input shadow quantities and output quantities. They include the first-order conditions from the dual shadow cost-minimization problem, computing firm- and input-specific AE parameters that are time-varying. Since input and output quantities are endogenous, instruments which over-identify the model are firm dummies, time dummies, and interactions of firm with time dummies. They find substantial allocative inefficiencies for capital and energy. The average RTS is 1.17 for their sample, which is similar to that obtained by Caves, Christensen, and Swanson (1981), who compute a variable multiple-output cost system using a pooled sample with more US railroads but fewer years. For the full 1955-74 period, they calculate an average annual rate of productivity growth of 1.8%.

A number of studies of railroads formulate production or factor requirement functions but fail to adequately address the potential endogeneity of inputs. Perelman and Pestieau (1988) estimate a production function for a pooled sample of 19 non-US railroads for 1970-83 and perform a second-stage regression to explain the sources of productivity growth. The caveats of WS apply to the second-stage regression. Further, Perelman and Pestieau do not address the potential endogeneity of inputs and outputs. Coelli and Perelman (1999) formulate input and output-oriented distance functions, pooling annual data on 17 EU railway companies observed from 1988 to 1993. They find increasing RTS at the
mean of their data. Their work is subject to the same concerns regarding the potential endogeneity of inputs and outputs. Gathon and Perelman (1992) estimate FE and RE factor requirements functions for 19 EU railway companies from 1961-88. They make strong assumptions about exogeneity that may be invalid. Friebel, Ivaldi, and Vibes (2010) examine the effects of reforms on railroad efficiency for 11 pooled EU countries for approximately 20 years. They estimate a Cobb-Douglas production frontier as a function of capital, labor, a deregulation dummy, and country FE, where the latter two are scaled by time. After controlling for endogeneity using the interaction of time with the firm dummy and a deregulation dummy, they find that efficiency levels increase due to market reforms. These include third-party network access, introduction of an independent regulator, and vertical separation. One must question whether their FE approach accounts for all potential endogeneity.

3.3 Airlines

Many studies investigate RTS and productivity growth in airlines using panel data. The consensus is that roughly constant RTS prevail, that productivity growth is positive but modest, that deregulation reduces costs, and that substantial differences in firm efficiencies exist. However, many studies use a two-step method to explain productivity growth while others fail to carefully address endogeneity.

Caves, Christensen, and Tretheway (1981) formulate a model of TFP for 11 trunk US airline carriers from 1972-77. In the first step they compute indices of the log of TFP and differences in the logs of TFP in successive years. In the second step they regress these measures on time and firm dummies, output, average stage length, and load factor. Clearly, one must consider the caveats of WS about the potential bias from two-step estimation.

Nearly all succeeding studies estimate a FE model, sometimes in conjunction with other models. Schmidt and Sickles (1984) formulate a stochastic frontier production function for 12 US airlines with quarterly data from 1971-78. Inputs are capital, labor, energy, and materials, while output is capacity ton miles. They compare the within, generalized least squares (GLS), and maximum likelihood (ML) estimators. For the GLS estimator, they assume the independence of the firm effects and the explanatory variables. For the ML estimator they assume independence of the firm effects and the idiosyncratic error term from themselves and from the regressors. Further, they assume a distribution for each component of the error. Estimated technical efficiencies are relatively constant across specifications. Productivity growth ranges from 1.5 to 2%. A Hausman (1978) test of the null hypothesis of no correlation between the effects and the regressors is accepted. However, the low power of this test implies a low probability of rejecting a false null. We must also consider the possibility of other sources of endogeneity not
eliminated by their FE approach.

Park, Sickles, and Simar (1988) specify airline production functions using a panel of US and European airlines from 1976-90 to explain capacity ton-mile service. They compute a within, semiparametric efficient IV estimator, where the random effects and the regressors have certain patterns of correlation. Their results indicate constant RTS in the provision of service capacity and that productivity growth is slightly more that 2%.

Cornwell, Schmidt, and Sickles (1990) consider efficient IV estimation of a frontier production function for ton miles using a panel of quarterly data on 8 US airlines from 1970-81. They allow coefficients and intercepts to vary over firms and measure time-varying technical efficiency levels for individual firms. Their approach avoids strong distributional assumptions for technical inefficiency and random noise by including in the production function a flexible function of time and firms. The authors compare the within estimator, a GLS estimator, and an extension of the Hausman and Taylor (1981) IV estimator, assuming that seasonal dummies, a time trend, labor, and materials are exogenous. Estimates from the within transformation, GLS, and efficient IV are very similar. The computed RTS are not significantly different from unity. Productivity growth rates average slightly greater than 1% for the first and third methods.

Good, Röller, and Sickles (1995) compute a FE estimator for a frontier production function using a panel of the eight largest EU and US airlines from 1976-86. While the US industry is deregulated, the EU industry not totally deregulated but has become increasingly more competitive. The authors employ many of the same variables use by Schmidt and Sickles (1984). Good et al. find that the American carriers are more productively efficient. However, endogeneity may be a problem with both studies.

Sickles, Good, and Johnson (1986) estimate a FE profit function for a panel of 13 US airlines observed quarterly from 1970-81. They obtain prices for passenger and cargo revenue ton-miles as well as capital, labor, energy, and materials inputs. Since the profit function includes service output characteristics (service quality and stage length), the authors specify reduced-form equations for them which are functions of prices and time. They formulate a system comprised of these equations, the profit function, and the output supply and input demand equations. Following the methods of Atkinson and Halvorsen (1980, 1984), they include parameters to measure allocative distortions. A firm-specific error-component term measures time-invariant unobserved heterogeneity. They conclude that deregulation reduces both the total cost of allocative distortions and their relative levels for input and outputs.

Two studies estimate FE cost systems. Atkinson and Cornwell (1994) formulate a firm FE shadow
cost system explaining capacity ton-miles for a panel of 13 US airlines using quarterly data from 1970-84. They measure firm-specific TE and calculate firm- and input-specific parameters measuring AE. Substantial inefficiencies exist. Potential cost savings from achieving TE and AE vary from 14 to 48% across airlines. Baltagi, Griffin, and Rich (1995) specify firm FE using a short-run cost system (capital is fixed) with panel data for 24 US airlines from 1971-86. They separate cost changes into categories attributable to technical change, economies of scale and density, and input prices. After a first-step estimation of a general index of industry TC, a second-step regression explains the sources of TC. They determine that deregulation stimulates technical change due to more efficient route structures. However, their two-step methodology is subject to the WS caveats.

Gerardi and Shapiro (2009) formulate FE and cross-sectional models using quarterly panel data on nine major US airline carriers from 1993-2006 to explain price dispersion. They determine that competition has a negative effect on price dispersion, which is the expected result. These results contrast with those of Borenstein and Rose (1994), who find that price dispersion increases with competition, based on a 10% random sample of US airfares for the second quarter of 1986. Using cross-sectional data, Gerardi and Shapiro basically reproduce the Borenstein-Rose results. However, after controlling for route-carrier characteristics, their FE estimator yields the opposite result. A reasonable conclusion is that this estimator eliminates the omitted variable bias induced by time-invariant, route-carrier effects. Unfortunately, few other studies carry out such a comparison, which would help greatly in determining the relative value of the two approaches.

A number of studies fail to employ instruments where endogeneity seems to be clearly problematic. Liu and Lynk (1999) formulate a variable cost function using a panel of 11 US airlines for the post-deregulation period 1984-91. They include stage length and load factor, which are arguably endogenous, but do not utilize instruments. While they fail to reject the null of exogeneity using a Hausman test, the power of the Hausman test is low, as mentioned previously. Sun (2017) investigates the impact of airline deregulation on air travel demand and competition using pooled panel data for three Korean routes from 2006-10. The explanatory variables for market shares are arguably endogenous. Whalen (2007) includes firm FE in regressions explaining price and the number of passengers using an 11 year panel data set of US and EU airlines. He investigates code sharing, antitrust immunity, and the open skies treaties. The use of lagged instruments may be problematic and the inclusion of a number of macro variables, such as population and income, make inference difficult.
4 A Control Function Approach

As an alternative to the estimation of firm FE, an important line of research specifies a two component error term that is different from that of (1). The first is an idiosyncratic shock that is a surprise to the firm and hence cannot be predicted or observed. The second is an error that is potentially observable or predictable productivity, which is at least in part known by the firm but is unknown to the econometrician. If the firm at least partially takes this error into account in its choice of inputs, it is correlated with the explanatory variables. As indicated above, GM argue that the within estimator distorts computed RTS, eliminates important identifying information, will not eliminate time-varying endogeneity, and increases measurement error. Alternatively, Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), and Ackerberg, Caves, and Frazer (2015) (ACF) consider a Cobb-Douglas panel-data model for firm \( i \) in period \( t \) that explicitly models the productivity component of the error term rather than resorting to time-demeaning of the data.

These three papers use a control function approach to model the productivity shock component of the error term. They introduce a monotonic function for productivity into a Cobb-Douglas production function:

\[
y_{it} = x_{it}\beta + \omega_{it} + e_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{6}
\]

where \( y_{it} \) is the dependent variable, \( x_{it} \) is a \((1 \times K)\) vector of regressors, \( \beta \) is a \((K \times 1)\) vector of coefficients, \( \omega_{it} \) is a productivity shock, and \( e_{it} \) is an idiosyncratic error term. The major contribution of this approach is to replace \( c_i \) in (1) with \( \omega_{it} \) and to create a proxy for \( \omega_{it} \) by introducing a new function, which for OP is the investment function and for LP is an intermediate input demand function, to control for the unobserved productivity shock. Assuming a monotonic relationship between investment or intermediate demand and \( \omega_{it} \), one can invert the investment or intermediate demand function and solve for \( \omega_{it} \) as an unknown function of observed variables. The resulting control function is then substituted for \( \omega_{it} \) in (6) before the model is estimated. Differing from OP and LP, ACF argue that labor is a deterministic function of the set of variables in the OP/LP procedures, upon which one must condition. Thus, ACF suggest inverting investment or intermediate demand functions that are conditional on the labor input. With OP/LP/ACF, \( \omega_{it} \) is not fixed over time, allowing calculation of time-varying partial effects of explanatory variables directly from the estimated control function.

The control function approach is not without its own potential problems. The analyst must specify
a functional form for \( \omega_{it} \) and specify moment conditions to identify the parameters of the model. This requires assumptions about the lagged dependence of productivity and the validity of additional instruments such as input prices. Additionally, as pointed out by GM, one assumes that the control function is completely specified and, before inversion of the investment function or the intermediate demand function, that \( \omega_{it} \) rather than its change from last period is an argument in these functions. If these assumptions are incorrect, the control function approach does not model endogeneity correctly.

Borrowing the control function approach of OP/LP/ACF, Atkinson, Primont, Tsionas (2018) use Gibbs sampling to estimate a technology-oriented directional distance function together with price equations derived from profit-maximization and cost-minimization models. They employ a balanced panel of 77 US utilities from 1988-2005. Their Bayesian approach allows joint estimation of a generalized control function, a second-order flexible production function, and shadow price equations, where the latter allow measurement of firm-specific price inefficiency. The authors also compute optimal firm-specific directions. Employing input prices as part of their instrument set, these authors reach four major conclusions: 1) the profit-maximization model is superior to its cost-minimization counterpart; 2) only weak support exists for the AJ effect; 3) mean productivity change is slightly less than 1%; and 4) the partial effects of productivity with respect to its arguments are largest for lagged productivity and energy prices.

Following OP/LP/ACF, Atkinson and Luo (2018) formulate a Cobb-Douglas production function for electricity generation which includes a control function for productivity, using a panel of the 80 largest and almost exclusively coal-fired US electric power plants from 1995-2005. They extend OP/LP/ACF by modeling the optimal control of a major pollutant, \( \text{SO}_2 \). Subject to a given level of electricity production, they compute the cost-minimizing solution for the firm, which must either purchase emission permits or directly control this pollutant. They include input prices and lagged inputs as instruments, calculating partial effects of inputs on productivity from the fitted control function.

Both ACF and GM question whether input and output prices are valid instruments. If the Law of One Price holds, variation in prices occurs only because of unobserved quality differences, rendering prices correlated with the error term, which contains these quality measures. However, with electric utilities, omitted quality differences are minimal. For wages, unmeasured quality differences should be important only for higher-level management. Other tasks are highly mechanized in a very capital-intensive process. The price of energy is in terms of thermal content, so there is no omitted quality differential. The price of capital is typically a function of the price index for equipment and structures, the yield on utility bonds, tax rates, the ratio of equity to total capitalization, and depreciation. This
calculation includes all important quality differentials. Finally, the price of output is measured as price per kilowatt hour in terms of a standardized voltage, which includes the relevant quality measure. To a large extent, similar arguments can be made for the use of non-labor input prices as instruments in the analysis of railroads and airlines.

What factors might cause exogenous input price variation across time and firms? Many exogenous market imperfections vary across firms and time: 1) the degree of union power which is greater in the Northeast, but which has diminished over time nationwide; 2) increased availability of natural gas in the Northeast, which historically receives less than the South, due to recent fracking; 3) recent state subsidies for the purchase of high-sulfur coal if made within same-state boundaries; and 4) increasingly strict state air quality implementation plans in states such as California, which affect the choice of fuel relative to pollution-control equipment.

5 Conclusion

A general consensus emerges for three important sectors of panel data energy economics. Substantial agreement exists among studies of the electric power industry, whether the unit of observation is the utility or plant and whether the data is cross-sectional or panel. With the latter, general agreement exists among studies that pool panel-data and those that estimate firm FE to remove time-invariant unobservables. A wide consensus of results is also found across a variety of behavioral assumptions: cost minimization, profit maximization, distance functions without behavioral assumptions, and systems of equations that append first-order conditions to cost, profit, and distance functions. Generally, studies reject homogeneous and homothetic functional forms and find that RTS are nearly constant for the largest firms. Further, productivity growth, computed as a residual, declines over time to small but positive levels. Nearly all panel studies examining vertical integration find substantial EVI, which argues against restructuring of the generation, transmission, and distribution sectors. Cost saving may accrue from a competitive generating sector. Substantial controversy remains regarding the existence of an AJ effect and the relative efficiency of public versus private utilities. Railroads appear to enjoy increasing RTS but exhibit substantial allocative inefficiency through 1975, with low productivity growth. Airlines appear to operate close to constant RTS, while TE and allocative inefficiency are substantial for many firms. Both decrease with deregulation. With renewed interest in production function estimation, the approach of OP/LP/ACF is a potentially valuable method for dealing with endogeneity by directly estimating the unknown productivity term using a control function. However, some of their identifying
assumptions may be problematic. Bayesian methods are an important tool to facilitate the estimation of more complex models of this variety. Unfortunately, many recent studies of production functions ignore much of this literature, assume a homogeneous technology, and ignore possible endogeneity, much like the initial production function literature for the electricity sector in the 1960s. The major difference is that the new studies employ panel rather than cross-sectional data.

References


