JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **15**: 187–210 (2000)

SCALE ECONOMIES IN ELECTRICITY DISTRIBUTION: A SEMIPARAMETRIC ANALYSIS

A. YATCHEW*
Department of Economics, University of Toronto, 150 St George Street, Toronto, Canada M5S 3G7

SUMMARY
We estimate the costs of distributing electricity using data on municipal electric utilities in Ontario, Canada for the period 1993–5. The data reveal substantial evidence of increasing returns to scale with minimum efficient scale being achieved by firms with about 20,000 customers. Larger firms exhibit constant or decreasing returns. Utilities which deliver additional services (such as water/sewage), have significantly lower costs, indicating the presence of economies of scope. Our basic specifications comprise semiparametric variants of the translog cost function where output enters non-parametrically and remaining variables (including their interactions with output) are parametric. We rely upon non-parametric differencing techniques and extend a previous differencing test of equality of non-parametric regression functions to a panel data setting. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

In various parts of the world electricity industries have been undergoing restructuring. The main driver has been a fundamental change in the economics of *generating* electricity. As a result of low natural gas prices and improved gas turbine technology, minimum efficient scale has fallen dramatically in this segment of the industry so that competitive markets in generation can be established. *Transmission* and *distribution* of electricity, however, continue to be natural monopolies. Ownership of such facilities conveys considerable market power and thus continues to attract regulatory oversight.

The structure and ownership of electricity industry varies. In many jurisdictions, all stages of the electricity production process — generation, transmission and distribution — are dominated by a single vertically integrated firm which may be privately or publicly owned. In others, distinct firms subsume varying combinations of these functions. For example, a number of jurisdictions have multiple generating companies some of which also own transmission and distribution. Only a few jurisdictions have many distribution companies.

In this paper we focus on the economics of *distributing* electricity. While there are many empirical studies which analyse the electricity industry as a whole or the generation segment specifically, there are precious few (we will reference them momentarily) which deal principally with distribution. There are three reasons for this. First, many regulatory jurisdictions have too few distinct entities engaged in distribution to permit serious statistical analysis. Second, inter-jurisdictional comparisons are difficult because of varying accounting practices and differing definitions of distribution (which is usually defined by the voltage at which power is taken from the transmission system). Third, where distribution is part of a vertically integrated utility — and this is frequently the case — a clear and comparable separation of distribution costs from those of other stages of production is typically not available.

* Correspondence to: Professor A. Yatchew, Department of Economics, University of Toronto, 150 St George Street, Toronto, Canada M5S 3G7; e-mail: yatchew@chass.utoronto.ca

Received 30 September 1997
Revised 11 May 1999

The data we examine do not suffer from these difficulties. Our analysis involves 81 municipal distributing utilities in Ontario, Canada, ranging in size from about 600 to 220,000 customers. Accounts are kept on a uniform basis as prescribed by the regulator, thus facilitating comparability in empirical work. Almost all power delivered by these utilities is purchased from Ontario Hydro, the dominant provincial generator — very few are involved in self-generation or ownership of transmission facilities.

Three recent studies examine electricity distribution costs in detail. Giles and Wyatt (1993) estimate a total cost function for 60 distributors in New Zealand. Salvanes and Tjotta (1994) estimate a variable cost function for 100 Norwegian distributors. Both of these studies are cross-sectional. Filippini (1996, 1997) estimates both variable and total cost functions using panel data on 39 Swiss distributors for the period 1987–91. All the above studies find evidence of scale economies in the distribution of electricity, though for New Zealand and Norway, minimum efficient scale occurs at surprisingly small levels of operation.[1]

In this paper we estimate total cost functions where total costs (*TC*) consist of operations and maintenance (*OM*), billing/collection/administration (*BCA*), depreciation (*DEP*) and interest (*INT*) costs. While each of the above three studies include the cost of power in their dependent variable, we exclude this component in order to focus exclusively on the distribution service provided by the utility. The cost of power, which includes generation and transmission, dominates distribution costs (in Ontario by a factor of more than 5 : 1), thus even small percentage errors in its measurement could substantially reduce the accuracy of estimates of distribution cost parameters. In Ontario, the prices paid for this power are established through a quasi-regulatory process. Volume discounts on power purchases are not available. Thus, one would not expect any significant scale economies in the power procurement function of these utilities, adding justification to our analysis of pure distribution costs. We find substantial evidence of increasing returns to scale with minimum efficient scale being achieved by firms with about 20,000 customers. Larger firms exhibit constant or decreasing returns. Utilities which deliver additional services (46% of the utilities under study provide other municipal services — such as water/sewage) have lower costs, indicating the presence of economies of scope.

Our basic specifications are variants of the translog cost function where output enters non-parametrically while remaining variables (including their interaction with output) enter in a parametric fashion. Section 2 describes the model and provides additional details about the data. Section 3 uses single-equation differencing techniques to analyse costs. Section 4 outlines differencing estimation of variance components, extends a previous differencing test of equality of non-parametric regression functions (Yatchew, 1998b) to this setting and reports results of tests applied to the panel data. For an alternative test procedure see Baltagi *et al.* (1996). The concluding Section 5 compares our results to those of previous studies.

2. MODEL AND DATA

During the period of our analysis there were approximately 300 municipal distributing utilities in Ontario. We use the 81 utilities for which the most complete data exist. Since data tend to be missing for small utilities, our truncated data-set actually represents over 70% of the municipal

[1] Other studies which focus on the distribution sector include Neuberg (1977), Nelson (1990), Hjalmarsson and Veiderpass (1992a,b) and Salvanes and Tjotta (1998). For studies of the generation segment or vertically integrated electric utilities see references contained in e.g. Pollitt (1995) or Kwoka (1997).

distributor customer base. In 1995, Ontario municipal distributors purchased about 90 twh of electricity at 6·5 ¢/kwh. (Throughout the paper, monetary amounts are in Canadian dollars.) The electricity was sold to end-users at an average price of about 7·6 ¢/kwh. Thus total municipal distributor revenues were about \$1 billion after subtracting the cost of power. Variable costs which consist of *OM* plus *BCA* costs represent about 52% of this amount (with *OM* and *BCA* costs taking 27% and 25% shares respectively). Depreciation and interest expense represent 25% and 3%, (collectively, these utilities have very low debt). The remaining 20% of revenues flows to net income.

Our main empirical objective is to estimate scale economies of delivering electricity. *A priori*, the relationship between firm size and unit costs may be flat, increasing, decreasing or U-shaped; it may be concave or it may have multiple inflection points. We propose therefore to estimate the scale effect using a semiparametric model.

In addition to the level of output, a number of variables may influence costs and therefore need to be incorporated into the model. These covariates include the conventional arguments of cost functions — the price of labour which we measure by the hourly wage (*WAGE*) of linemen of identical grade; and the price of capital (*PCAP*), which we measure by dividing accumulated gross investment in plant and facilities (*TOTPLANT*) by total kilometres of distribution wire (*KMWIRE*).

We also include a series of covariates which reflect differences among utilities. Since the level of service to a 'typical' customer will in general influence costs, we include the total quantity of electricity delivered per customer (*KWH/CUST*). The remaining lifetime of assets (*LIFE*) is included to allow for vintage effects — for example, one might expect older capital to require more maintenance. Load factor (*LF*) — which measures capacity utilization relative to peak usage — is included since high load factor utilities require greater expenditures in order to maintain reliability. There is considerable variation in the density of customers across utilities. To capture this effect, we divide the total kilometres of distribution wire by the number of customers (*KMWIRE/CUST*). One would expect higher costs, the greater the distance between customers.

About 46% of our distributors (37 of 81) are part of local Public Utility Commissions (PUCs) which deliver additional services such as water and sewage removal. The regulator requires that costs of the various services be separated as far as possible and provides detailed accounting rules for this purpose (see Ontario Hydro, 1995). Although operations are indeed separate, other functions (e.g. billing and collection) are performed on a shared basis and each service is allocated a pro rata share of costs. Thus, one would expect PUCs to exhibit some cost savings. One of our objectives will be to assess whether this is indeed the case.

Our basic econometric specification is given by:

$$\begin{aligned}\log(TC/CUST) = {} & f(\log(CUST)) \\ & + \beta_1 \log(WAGE) + \beta_2 \log(PCAP) \\ & + \tfrac{1}{2}\beta_{11} \log^2(WAGE) + \tfrac{1}{2}\beta_{22} \log^2(PCAP) + \beta_{12} \log(WAGE)\log(PCAP) \\ & + \beta_{31} \log(CUST)\log(WAGE) + \beta_{32} \log(CUST)\log(PCAP) \\ & + \beta_4 PUC + \beta_5 \log(KWH/CUST) + \beta_6 \log(LIFE) + \beta_7 \log(LF) \\ & + \beta_8 \log(KMWIRE/CUST) + v\end{aligned} \tag{1}$$

We assume little about the function f beyond smoothness, thus, equation (1) is a translog cost function, with the output variable (*CUST*) entering both non-parametrically (through f) and parametrically (through the interaction terms between output and the price variables). It is readily verified that if these interaction terms are zero (i.e. $\beta_{31} = \beta_{32} = 0$) then the cost function is homothetic. The model has a partial linear structure $y = f(x) + z\beta + v$ where the *non-parametric* variable x is $\log(CUST)$ and the vector z is composed of the various price and other variables which enter parametrically. We adopt the notational convention that lower-case italicized names represent transformed variables in logarithmic form, e.g. $kwh = \log(KWH/CUST)$. For convenient reference variable definitions and summary statistics are contained in Appendix 2.

The partial linear structure is amenable to some particularly simple 'differencing' techniques because the parametric and non-parametric portions of the model are additively separable. The essential idea is to reorder the data so that the values of the non-parametric variable are 'close', then to take first- or higher-order differences to remove the non-parametric effect. We will avail ourselves of this device extensively in this paper.[2]

When applying the differencing procedures used in this paper, the first few observations may be treated differently or lost.[3] For the mathematical arguments below, such effects are negligible. Thus, we will use the symbol $\doteq$ to denote 'equal except for end effects'. In the panel data portion of the paper, asymptotics are on N (the number of observations in each period) with T (the number of time periods) fixed. Throughout, I_N, I_T, I_{NT} will denote identity matrices of dimension N, T and $N * T$ respectively, ι_T will be a $T \times 1$ column vector of ones, $1_T = \iota_T \iota_T'$ a T-dimensional square matrix of ones. We use the abbreviations *tr* for trace, *dim* for dimension. $A \odot B$ is the matrix whose *ij*th entry is $A_{ij}B_{ij}$.

3. SINGLE-EQUATION ANALYSIS

3.1 Basic Setup and Differencing Procedures

Our model may be written in the form:

$$y_{it} = f_t(x_{it}) + z_{it}\beta_t + v_{it} \tag{2}$$

where $t = 1, 2, 3$ for the years 1993–5 and $i = 1, \ldots, N$ indexes utilities. Throughout the paper, the non-parametric variable x_{it} is a scalar.

Let $y_t = (y_{1t}, \ldots, y_{Nt})'$ be the N-dimensional column vector of the values of the dependent variable in year t. Define $x_t = (x_{1t}, \ldots, x_{Nt})'$ and $v_t = (v_{1t}, \ldots, v_{Nt})'$ in a similar fashion. We assume that for fixed t, the residuals are distributed independently and homoscedastically across firms. For each firm, the k-dimensional row vector z_{it} contains data on the parametric variables and we define the $N \times k$ matrix $Z_t = (z_{1t}', \ldots, z_{Nt}')'$. We emphasize that for purposes of this section, the data have already been ordered so that *within* each year, the x's are in increasing order, i.e. $x_{1t} \leqslant \cdots \leqslant x_{Nt}$, $t = 1, 2, 3$. In matrix notation, we write our model as:

$$y_t = f_t(x_t) + Z_t\beta_t + v_t \tag{3}$$

where $f_t(x_t)' = (f_t(x_{1t}), \ldots, f_t(x_{Nt}))$.

[2] Differencing has been used in the pure non-parametric regression model by Rice (1984), Hall *et al.* (1990), Yatchew (1988, 1998b) and Cox and Koh (1989). In the partial linear model it has been used by Powell (1987), Ahn and Powell (1993) and Yatchew (1997).

[3] For example, the simplest differencing estimator of the residual variance in a non-parametric regression is of the form $s^2 = \Sigma_i^N (y_i - y_{i-1})^2/2N$, whence the summation begins at $i = 2$.

Let m be the order of differencing and $d_0, d_1, \ldots, d_m$ the optimal differencing weights.[4] The weights satisfy the conditions:

$$\sum_{j=0}^{m} d_j = 0 \quad \sum_{j=0}^{m} d_j^2 = 1 \tag{4}$$

The first condition ensures that differencing removes the non-parametric effect as sample size increases and the x's become close, (see equation (6) below). The second condition is a normalization which implies that the residuals in the differenced equation (6) have the same variance as those in the original equation (3). Define the differencing matrix:

$$\underset{N\times N}{D} = \begin{bmatrix} d_0, d_1, d_2, \ldots, d_m, 0, \ldots\ldots\ldots\ldots, 0 \\ 0, d_0, d_1, d_2, \ldots, d_m, 0, \ldots\ldots\ldots, 0 \\ \vdots \quad \vdots \\ \vdots \quad \vdots \\ 0, \ldots\ldots\ldots\ldots, 0, d_0, d_1, d_2, \ldots, d_m, 0 \\ 0, \ldots\ldots\ldots\ldots\ldots, 0, d_0, d_1, d_2, \ldots, d_m \\ 0, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, 0 \\ \vdots \quad \vdots \\ 0, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, 0 \end{bmatrix} \tag{5}$$

(Properties of D and related matrices are summarized in Appendix 1.) Application of the differencing matrix to model (3) permits direct estimation of the parametric effect. In particular, take:

$$Dy_t = Df_t(x_t) + DZ_t\beta_t + Dv_t \tag{6}$$

Since the data have been reordered so that the x's are close, the application of the differencing matrix D in model (6) removes the non-parametric effect in large samples. Under general conditions, the OLS regression of Dy_t on DZ_t exhibits the following large sample behaviour (see Appendix 1 for further details):

$$\hat{\beta}_t = [(DZ_t)'DZ_t]^{-1}(DZ_t)'Dy_t \overset{A}{\sim} N\left(\beta_t, \left(1 + \frac{1}{2m}\right)\frac{\sigma_v^2}{N}\Sigma_{z|x}^{-1}\right) \tag{7a}$$

where $\Sigma_{z|x} = E[\mathrm{Cov}(z|x)]$ is estimated consistently using

$$\hat{\Sigma}_{z|x} = \frac{1}{N}(DZ_t)'DZ_t \tag{7b}$$

[4] Hall *et al.* (1990) propose optimal weights for non-parametric differencing procedures. They claim that optimal differencing weights are unique. In fact, as m increases, there is an increasing number of optimal differencing sequences but only one which corresponds to an MA process with roots on or outside the unit circle. For the unique optimal differencing weights satisfying the latter property see Yatchew (1998a, p. 697).

the residual variance is estimated consistently using

$$s_v^2 = \frac{1}{N}(Dy_t - DZ_t\hat{\beta}_t)'(Dy_t - DZ_t\hat{\beta}_t) \tag{7c}$$

and the covariance matrix of the differencing estimator of β may be estimated using:

$$\hat{\Sigma}_{\hat{\beta}_t} = \left(1 + \frac{1}{2m}\right)\frac{s_v^2}{N}\hat{\Sigma}_{z|x}^{-1} \tag{7d}$$

By increasing the order of differencing m, the estimator may be shown to be asymptotically efficient. In all applications of the differencing estimator, the principal requirement is that the average distance between the values of the non-parametric variable x decline to zero sufficiently quickly. (For details, see Yatchew, 1997.)

Linear restrictions of the form $R\beta_t = r$ may be tested using the conventional statistic $(R\hat{\beta}_t - r)'(R\hat{\Sigma}_{\hat{\beta}_t}R')^{-1}(R\hat{\beta}_t - r)$ which converges in distribution to a chi-square with degrees of freedom equal to the rank of R.

3.2 Discussion of Empirical Results

Differencing estimates of the parametric component of the Full Model, equation (1), are presented for the years 1993 to 1995 in Tables I(a)–(Ic). (Throughout the paper we use third-order differencing ($m = 3$). Results for other orders of differencing were similar.) We do not find significant statistical evidence against either the Homothetic Model or the Loglinear Models. Focusing on the latter, the estimated *wage* effect is positive and moderately significant while the effect of *pcap* is positive and strongly significant. The estimate of cost savings associated with distributors that are part of a Public Utility Commission ranges from 7% to 10%. The level of per

Table I(a). Semiparametric analysis of total costs — 1993

	Full Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
wage	1·6303	13·883	1·5082	12·999	0·3543	0·3119
pcap	−3·5079	2·5194	−2·0913	1·5894	0·5041	0·0676
$\frac{1}{2}wage^2$	−3·7813	4·9276	−2·2686	4·2613	—	—
$\frac{1}{2}pcap^2$	0·1667	0·1636	0·0954	0·1320	—	—
$wage \cdot pcap$	0·7795	0·7067	0·4867	0·5791	—	—
$cust \cdot wage$	0·1298	0·1787	—	—	—	—
$cust \cdot pcap$	−0·0351	0·0479	—	—	—	—
PUC	−0·0855	0·0386	−0·0893	0·0384	−0·0870	0·0378
kwh	0·0476	0·0841	0·0476	0·0841	0·0301	0·0838
life	−0·6002	0·1168	−0·6099	0·1151	−0·6265	0·1144
lf	0·5047	0·2256	0·5789	0·2036	0·6016	0·2019
kmwire	0·3573	0·0856	0·3593	0·0860	0·3690	0·0849
s_v^2	0·0184		0·0185		0·0193	
R^2	0·666		0·664		0·650	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 0·54. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 2·77. Order of differencing $m = 3$.

Table I(b). Semiparametric analysis of total costs — 1994

	Full Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
wage	−1·0949	13·4211	−1·0005	12·4863	0·4726	0·2870
pcap	−1·6346	2·2463	−1·1040	1·4690	0·6176	0·0608
$\frac{1}{2}wage^2$	−1·2183	4·7920	−0·6763	4·2536	—	—
$\frac{1}{2}pcap^2$	0·0927	0·1433	0·0687	0·1205	—	—
$wage \cdot pcap$	0·4182	0·6630	0·3003	0·5457	—	—
$cust \cdot wage$	0·0457	0·1512	—	—	—	—
$cust \cdot pcap$	−0·0127	0·0407	—	—	—	—
PUC	−0·1053	0·0349	−0·1062	0·0348	−0·1029	0·0335
kwh	0·0911	0·0739	0·0904	0·0735	0·0806	0·0718
life	−0·4848	0·0970	−0·4879	0·0955	−0·4930	0·0938
lf	0·3417	0·2040	0·3695	0·1834	0·3911	0·1795
kmwire	0·5433	0·0783	0·5439	0·0784	0·5485	0·0773
s_v^2	0·0149		0·0149		0·0152	
R^2	0·745		0·744		0·739	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 0·54. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 1·42. Order of differencing $m = 3$.

Table I(c). Semiparametric analysis of total costs — 1995

	Full Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
wage	8·4377	14·5471	4·2684	13·6000	0·5745	0·2692
pcap	0·2941	2·2338	0·0118	1·4807	0·5012	0·0583
$\frac{1}{2}wage^2$	−2·1468	5·0816	−0·9708	4·5408	—	—
$\frac{1}{2}pcap^2$	0·0588	0·1398	0·0593	0·1173	—	—
$wage \cdot pcap$	−0·1476	0·6621	−0·0648	0·5520	—	—
$cust \cdot wage$	0·0438	0·1491	—	—	—	—
$cust \cdot pcap$	−0·0032	0·0398	—	—	—	—
PUC	−0·0626	0·0349	−0·0649	0·0348	−0·0678	0·0329
kwh	0·0759	0·0724	0·0808	0·0720	0·0839	0·0689
life	−0·3397	0·0883	−0·3494	0·0875	−0·3584	0·0860
lf	0·3778	0·2040	0·3991	0·1851	0·3906	0·1788
kmwire	0·3850	0·0768	0·3836	0·0773	0·3798	0·0761
s_v^2	0·0149		0·0150		0·0151	
R^2	0·678		0·675		0·673	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 2·14. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 0·34. Order of differencing $m = 3$.

customer electricity sales (*kwh*) has a small and insignificant impact on costs. The remaining life of assets (*life*) has a strong impact on costs — firms with older plant experience substantially higher costs. Higher load factors (*lf*) evidently also result in significantly higher costs. Utilities with lower density and hence greater distances between customers (*kmwire*) have significantly higher costs. We note that estimates of non-price covariate effects exhibit little variation as one moves from the Full Model to the Homothetic and Loglinear Models.

For comparison purposes we provide estimates of the parametric analogues of the models in Tables II(a)–(c). These are the Translog, the Homothetic and the Loglinear Models where the level of output *cust* (=log(*CUST*)) is modelled using a quadratic. Estimates of price effects differ substantially between parametric and semiparametric versions of the Full and Homothetic

Table II(a). Parametric analysis of total costs — 1993

	Full Translog Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
cust	−5·3264	1·9171	−0·7776	0·1945	−0·8567	0·1672
$cust^2$	−0·0269	0·0477	0·0776	0·0212	0·0860	0·0178
wage	52·6179	24·1400	3·0115	13·5428	0·6407	0·3037
pcap	1·7469	3·0096	−1·3122	1·6537	0·5292	0·0686
$\frac{1}{2}wage^2$	−19·3633	7·8971	−2·5065	4·4294	—	—
$\frac{1}{2}pcap^2$	−0·0024	0·1641	0·0411	0·1351	—	—
$wage \cdot pcap$	−0·5685	0·8586	0·4449	0·5762	—	—
$cust \cdot wage$	1·5803	0·6246	—	—	—	—
$cust \cdot pcap$	0·0543	0·0584	—	—	—	—
PUC	−0·0859	0·0370	−0·0841	0·0381	−0·0821	0·0366
kwh	0·0239	0·0825	0·0152	0·0856	0·0020	0·0828
life	−0·5147	0·1166	−0·6036	0·1154	−0·6097	0·1124
lf	0·2737	0·2359	0·5735	0·2044	0·5742	0·2009
kmwire	0·3196	0·0856	0·3915	0·0836	0·3989	0·0814
s_v^2	0·0194		0·0210		0·0214	
R^2	0·647		0·620		0·613	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 6·43. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 0·32.

Table II(b). Parametric analysis of total costs — 1994

	Full Translog Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
cust	−2·6463	1·7335	−0·6746	0·1785	−0·7226	0·1546
$cust^2$	0·0229	0·0414	0·0644	0·0195	0·0700	0·0164
wage	18·7046	23·4773	−3·9877	13·6240	0·7541	0·2875
pcap	1·1019	2·7553	−0·1492	1·5490	0·6270	0·0637
$\frac{1}{2}wage^2$	−6·8931	7·5552	0·6182	4·6111	—	—
$\frac{1}{2}pcap^2$	0·0012	0·1517	0·0015	0·1277	—	—
$wage \cdot pcap$	−0·2313	0·8290	0·2408	0·5446	—	—
$cust \cdot wage$	0·6800	0·5496	—	—	—	—
$cust \cdot pcap$	0·0214	0·0532	—	—	—	—
PUC	−0·1102	0·0357	−0·1088	0·0358	−0·1028	0·0337
kwh	0·0430	0·0763	0·0429	0·0770	0·0309	0·0734
life	−0·4354	0·1056	−0·4856	0·0987	−0·4777	0·0954
lf	0·2725	0·2154	0·3948	0·1894	0·4127	0·1845
kmwire	0·5233	0·0813	0·5499	0·0790	0·5526	0·0767
s_v^2	0·0176		0·0179		0·0180	
R^2	0·698		0·693		0·691	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 1·56. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 0·02.

Table II(c). Parametric analysis of total costs — 1995

	Full Translog Model		Homothetic Model		Loglinear Model	
Variable	Coef	SE	Coef	SE	Coef	SE
cust	−1·5286	1·4527	−0·6918	0·1750	−0·6802	0·1495
$cust^2$	0·0522	0·0336	0·0670	0·0191	0·0656	0·0159
wage	13·1980	21·7723	2·1635	14·3457	0·7147	0·2721
pcap	0·2027	2·4039	0·5519	1·5228	0·5067	0·0606
$\frac{1}{2}wage^2$	−5·4914	7·3342	−0·9339	4·8095	—	—
$\frac{1}{2}pcap^2$	0·0092	0·1451	−0·0353	0·1226	—	—
$wage \cdot pcap$	0·1020	0·7190	0·1186	0·5373	—	—
$cust \cdot wage$	0·3696	0·4807	—	—	—	—
$cust \cdot pcap$	−0·0148	0·0449	—	—	—	—
PUC	−0·0781	0·0356	−0·0814	0·0356	−0·0814	0·0328
kwh	0·0566	0·0741	0·0620	0·0736	0·0624	0·0693
life	−0·3586	0·0915	−0·3798	0·0889	−0·3782	0·0867
lf	0·2927	0·2135	0·3888	0·1877	0·3860	0·1820
kmwire	0·3907	0·0779	0·4015	0·0769	0·4067	0·0752
s_v^2	0·0173		0·0175		0·0176	
R^2	0·625		0·621		0·620	

Test of Homothetic Model versus Full Model, χ_2^2 under H_0: 0·87. Test of Loglinear Model versus Homothetic Model, χ_3^2 under H_0: 0·04.

Models. This is perhaps not surprising given the low precision with which they are estimated. However, estimates of non-price covariate effects are similar. The R^2, which we define as $R^2 = 1 - s_v^2/s_y^2$ is 2–5% higher in the semiparametric specifications relative to the pure parametric ones.

Returning to our semiparametric specification, we may now remove the estimated parametric effect from the dependent variable and analyse the non-parametric effect. In particular, for purposes of the tests below, the approximation $y_{it} - z_{it}\hat{\beta}_t = z_{it}(\beta_t - \hat{\beta}_t) + f_t(x_{it}) + v_{it} \cong f_t(x_{it}) + v_{it}$ does not alter the large sample properties of the procedures.

We use the estimates of the Loglinear Model to remove the parametric effect. Figure 1 displays the ordered pairs $(y_{it} - z_{it}\hat{\beta}_t, x_{it})$ as well as kernel estimates of f_t bordered by 95% uniform confidence bands. Quadratic estimates of scale effects are also illustrated. Parametric null hypotheses may be tested against non-parametric alternatives using the statistic:

$$(mN)^{1/2} \frac{s_{res}^2 - s_v^2}{s_v^2} \xrightarrow{D} N(0, 1) \text{ under } H_0 \quad (8)$$

where s_{res}^2 is the estimate of the residual variance from the parametric regression and s_v^2 is the differencing estimate from model (7c) above.[5] (For details see Yatchew, 1997, Proposition 2.) If we insert a constant function for f then the procedure constitutes a test of significance of the scale variable x against a non-parametric alternative. The resulting statistics range from 8·15 to 10·09 indicating a strong effect of output on unit costs, that is, a strong scale effect. Next we test a quadratic model for output. The resulting test statistics vary from 1·68 to 2·86, suggesting that the quadratic model is likely inadequate even though the quadratic estimates lie within the

[5] Statistic (8) may also be used to test non-parametric null hypotheses such as monotonicity or concavity. In these cases, s_{res}^2 may be obtained from an isotonic regression.

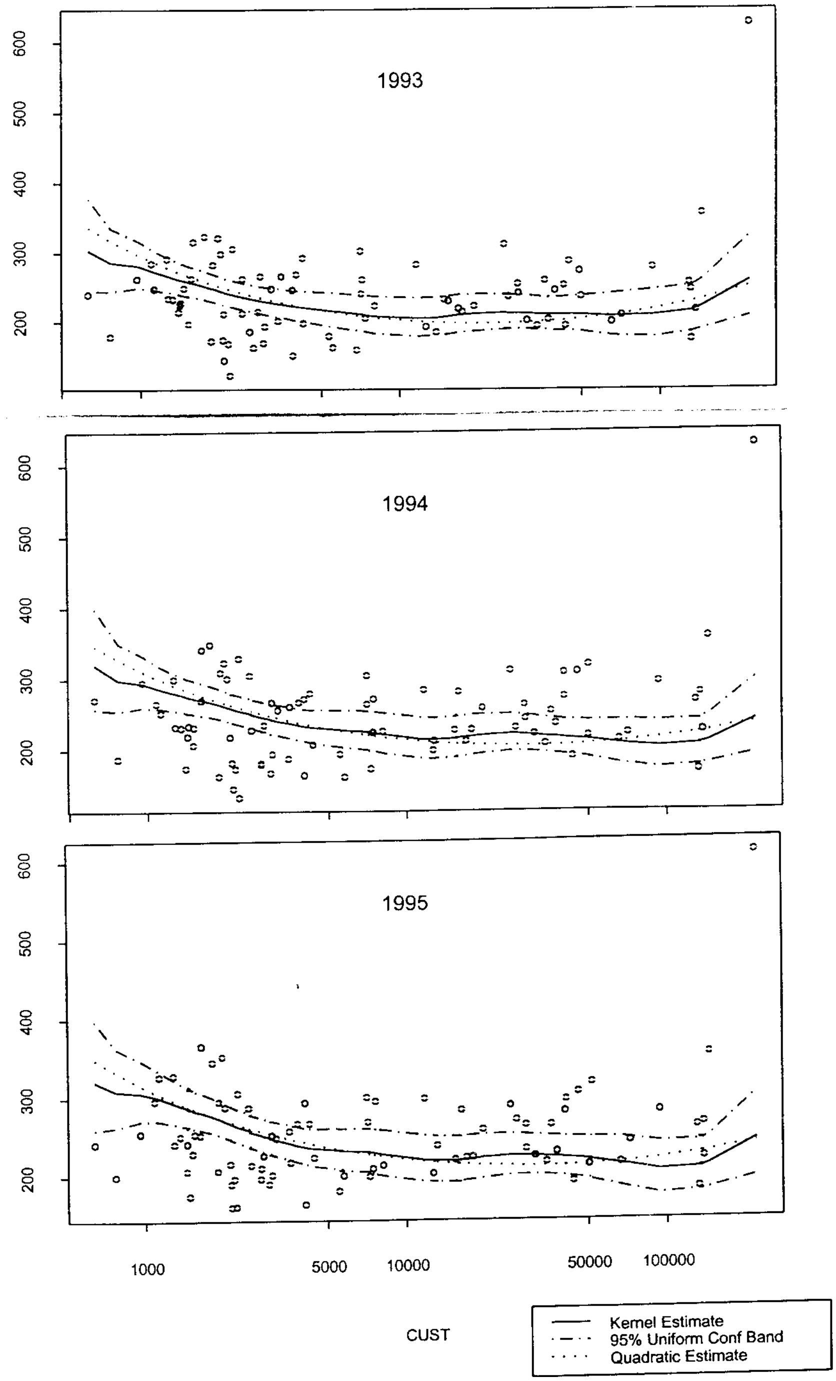

Figure 1. Single-equation analysis of total cost data — non-parametric component

asymptotic 95% confidence bands. Keeping in mind that (8) is a one-sided test, one would reject the parametric translog in favour of the semiparametric translog at the 5% level in each of the three years.

4. PANEL DATA ANALYSIS

4.1 Basic Setup

The availability of several years of data permits us to assess the stability of parametric effects over time as well as the stability of the non-parametric scale effect. The testing of these hypotheses will be the two main objectives of our panel data analysis. Recall that our basic model is given by $y_{it} = z_{it}\beta_t + f_t(x_{it}) + v_{it}$. We now elaborate the assumptions about the residual:

$$v_{it} = u_i + \varepsilon_{it} \tag{9}$$

where, conditional on the x's, $\mathrm{E}(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma_u^2$, $\mathrm{E}(u_i^4) = \eta_u$, $\mathrm{E}(\varepsilon_{it}) = 0$, $\mathrm{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$, $E\varepsilon_{it}^4 = \eta_\varepsilon$, $\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $s \neq t$, $\mathrm{Cov}(u_i, \varepsilon_{it}) = 0$ for all t. Define $u_0 = (u_1, \ldots, u_N)'$ — the subscript 0 is intended to connote that individuals are endowed with these effects at birth. We also assume that $((y_{i1}, x_{i1}), \ldots, (y_{iT}, x_{iT}))$ is independent of $((y_{j1}, x_{j1}), \ldots, (y_{jT}, x_{jT}))$ for $i \neq j$.

As before, let $v_t = (v_{1t}, \ldots, v_{Nt})'$ be the N-dimensional column vector of residuals during period t and $v = (v_1', \ldots, v_T')'$ the NT-dimensional concatenation of these column vectors. Define ε, x and y in a similar way. Then $v = \varepsilon + \iota_T \otimes u_0$ and $\mathrm{Cov}(v) = \sigma_\varepsilon^2 I_{NT} + \sigma_u^2(1_T \otimes I_N)$. Given T years of data, we write our model as:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} f_1(x_1) \\ f_2(x_2) \\ \vdots \\ f_T(x_T) \end{pmatrix} + \begin{pmatrix} Z_1 & 0 & 0 & \ldots & 0 \\ 0 & Z_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & Z_T \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{pmatrix} \tag{10}$$

$$y \quad = \quad f(x) \quad + \quad Z\beta \quad + \quad v$$

where, as before, $f_t(x_t)' = (f_t(x_{1t}), \ldots, f_t(x_{Nt}))$.

Unfortunately, the presence of individual effects will require us to carefully keep track of how data have been reordered. By convention, we will assume that in period 1, data are already ordered so that the x's are in increasing order. Data in all subsequent periods are *initially* in the same order as the data in the first period. This, of course, does not ensure that their corresponding x's are 'close', but it does ensure that the corresponding individual effects are in the same position in each year. We will need permutation matrices to reorder data and quadratic forms to estimate variances. To denote permutation matrices we will use P usually with a subscript and Q will denote matrices used in the construction of quadratic forms.

For each period t, define P_t to be the $N \times N$ permutation matrix which reorders the data within the period so that x's are in increasing order. Our above convention implies P_1 is the identity matrix. Define P_w, the $NT \times NT$ 'within' permutation matrix, to be the block diagonal permutation matrix with diagonal blocks $P_1, \ldots, P_T$. When applied to data stacked across all periods, it reorders so that corresponding x's are close within each period. Thus if $x^* = P_w x$ then $x_{1t}^* \leqslant \cdots \leqslant x_{Nt}^*$ for each t. Define P_p the 'pooled' permutation matrix to be the matrix which reorders data so that the x's are close regardless of which period they are in.

We transform the stacked model (10) by applying $(I_T \otimes D)P_w$ which yields:

$$\begin{pmatrix} DP_1 y_1 \\ DP_2 y_2 \\ \vdots \\ DP_T y_T \end{pmatrix} = \begin{pmatrix} DP_1 f_1(x_1) \\ DP_2 f_2(x_2) \\ \vdots \\ DP_T f_T(x_T) \end{pmatrix} + \begin{pmatrix} DP_1 Z_1 & 0 & 0 & \dots & 0 \\ 0 & DP_2 Z_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & DP_T Z_T \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix} + \begin{pmatrix} DP_1 v_1 \\ DP_2 v_2 \\ \vdots \\ DP_T v_T \end{pmatrix}$$

$$(I_T \otimes D)P_w y = (I_T \otimes D)P_w f(x) + (I_T \otimes D)P_w Z\beta + (I_T \otimes D)P_w v \quad (11)$$

The OLS estimator applied to these reordered, differenced and stacked data:

$$\hat{\beta} = (Z'P_w'(I_T \otimes D'D)P_w Z)^{-1} Z'P_w'(I_T \otimes D'D)P_w y \quad (12)$$

is identical to the estimator in equation (7a) applied equation by equation. However, its asymptotic covariance matrix must account for correlations between residuals over time arising out of the individual specific effect:

$$\begin{aligned} \Sigma_{\hat{\beta}} = {} & [Z'P_w'(I_T \otimes D'D)P_w Z]^{-1} Z'P_w'(I_T \otimes D'D)P_w \\ & \cdot [\sigma_\varepsilon^2 I_{NT} + \sigma_u^2 (I_T \otimes I_N)] \cdot P_w'(I_T \otimes D'D)P_w Z[Z'P_w'(I_T \otimes D'D)P_w Z]^{-1} \end{aligned} \quad (13)$$

and thus requires consistent estimation of σ_u^2 and σ_ε^2. We will need an estimate of $\Sigma_{\hat{\beta}}$ to perform tests on the parametric component of the model. Estimates of σ_u^2 and σ_ε^2 will also be used to test the stability of the non-parametric effect.

4.2 Estimation of Variance Components and a Test of Equality of Regression Functions

To simplify exposition, suppose that the parametric effect has been removed from the dependent variable, so that equation (10) becomes:

$$y = f(x) + v \quad (14)$$

Applying the 'within' permutation matrix and differencing yields:

$$(I_T \otimes D)P_w y = (I_T \otimes D)P_w f(x) + (I_T \otimes D)P_w v \quad (15)$$

To estimate $\sigma_v^2 = \sigma_u^2 + \sigma_\varepsilon^2$ define $Q_v = P_w'(I_T \otimes D'D)P_w$ and

$$s_v^2 = \frac{1}{NT} y' Q_v y \quad (16)$$

To estimate σ_u^2 define $Q_u = P_w'(I_T \otimes D')P_w((1_T - I_T) \otimes I_N)P_w'(I_T \otimes D)P_w$ and suppose $\hat{\pi}_u \xrightarrow{P} \pi_u > 0$, where

$$\hat{\pi}_u = \frac{1}{NT(T-1)} \operatorname{tr}(\iota_T \otimes I_N)' Q_u (\iota_T \otimes I_N)$$

Let

$$s_u^2 = \frac{1}{\mathrm{tr}(\iota_T \otimes I_N)' Q_u (\iota_T \otimes I_N)} y' Q_u y = \frac{1}{NT(T-1)} \frac{1}{\hat{\pi}_u} y' Q_u y \tag{17}$$

Reordering within periods and applying the differencing estimator removes the regression effect in large samples so that we have $(I_T \otimes D)P_w y \cong (I_T \otimes D)P_w v$. Thus the quadratic forms used to estimate σ_v^2 and σ_u^2 are approximately $v'Q_v v$ and $v'Q_u v$ respectively. To gain some intuition on how these quadratic forms yield estimates of the corresponding variances, suppose we are using first-order differencing. A typical term in $v'Q_v v$ will be of the form $\frac{1}{2}(v_{it}^* - v_{i-1t}^*)^2 = \frac{1}{2}(u_i^* + \varepsilon_{it}^* - u_{i-1}^* - \varepsilon_{i-1t}^*)^2$ the expectation of which is $\sigma_v^2 = \sigma_u^2 + \sigma_\varepsilon^2$. (Asterisks denote reordered data.)

Consider now the quadratic form $v'Q_u v$. After reordering and differencing, (that is, after applying $(I_T \otimes D)P_w$), we apply P_w', the inverse of the reordering matrix P_w. This realigns the data so that within each period, the ith firm forms part of the (differenced) ith observation. Interposing the matrix $(1_T - I_T) \otimes I_N$ takes covariances of differenced residuals across periods. For first-order differencing, a typical term in $v'Q_u v$ is given by $\frac{1}{2}(v_{is} - v_{i-1s}^*)(v_{it} - v_{i-1t}^{**}) = \frac{1}{2}(u_i + \varepsilon_{is} - u_{i-1}^* - \varepsilon_{i-1s}^*)(u_i + \varepsilon_{it} - u_{i-1}^{**} - \varepsilon_{i-1t}^{**})$ where single and double asterisks indicate reordering within periods s and t respectively. Thus the expectation of a typical term is $\frac{1}{2}(\sigma_u^2 + Eu_{i-1}^* u_{i-1}^{**})$ which equals σ_u^2 if the same firm precedes the ith firm in each of the two reorderings and $\frac{1}{2}\sigma_u^2$ otherwise. Dividing by $\mathrm{tr}(\iota_T \otimes I_N)' Q_u (\iota_T \otimes I_N)$ ensures that these consequences of reordering are properly taken into account when estimating σ_u^2.

For purposes of testing equality of regression functions, we will also want to use a statistic based on the pooled reordered data. Let $Q_p = P_p'(I_T \otimes D'D)P_p$ and define:

$$s_p^2 = \frac{1}{NT} y' Q_p y \tag{18}$$

Proposition 1 establishes consistency of s_v^2, s_u^2 and s_p^2. Proofs may be found in Appendix 1.

Proposition 1: (a) $s_v^2 \xrightarrow{P} \sigma_v^2$; (b) Suppose $\hat{\pi}_u \xrightarrow{P} \pi_u > 0$, then $s_u^2 \xrightarrow{P} \sigma_u^2$; (c) Suppose $\hat{\pi}_p = \mathrm{tr}(\iota_T' \otimes I_N) Q_p (\iota_T \otimes I_N)/NT \xrightarrow{P} \pi_p$ and all regression functions are identical, then $s_p^2 \xrightarrow{P} \sigma_\varepsilon^2 + \pi_p \sigma_u^2$. ■

Proposition 1 implies consistent estimation of $\sigma_\varepsilon^2 = \sigma_v^2 - \sigma_u^2$. In order to construct a test of equality of regression functions we will also need consistent estimates of the fourth-order moments η_u, η_ε.

Proposition 2: Let $d_0 = 1/\sqrt{2}$, $d_1 = -1/\sqrt{2}$ be the usual first-order differencing weights and D the corresponding first differencing matrix. Define

$$\hat{\eta}_u = \left(\frac{2}{NT(T-1)} \sum_{s \neq t} (P_s' D P_s y_s \odot P_s' D P_s y_s)'(P_t' D P_t y_t \odot P_t' D P_t y_t) - 4 s_u^2 s_\varepsilon^2 - 3\hat{\pi}_u s_u^4 - 2 s_\varepsilon^4 \right) \Big/ \hat{\pi}_u$$

$$\hat{\eta}_\varepsilon = \left(\frac{2}{NT} \sum_t (P_t' D P_t y_t \odot P_t' D P_t y_t)'(P_t' D P_t y_t \odot P_t' D P_t y_t) - \hat{\eta}_u - 12 s_u^2 s_\varepsilon^2 - 3 s_u^4 - 3 s_\varepsilon^4 \right) \tag{19}$$

where $P_1, \ldots, P_T$ are the $N \times N$ permutation matrices that make up the diagonal blocks of P_w. Then $\hat{\eta}_u \xrightarrow{P} \eta_u$ and $\hat{\eta}_\varepsilon \xrightarrow{P} \eta_\varepsilon$. ■

Proposition 3: Define

$$Q_\Upsilon = Q_p - Q_v + \frac{1-\hat{\pi}_p}{(T-1)\hat{\pi}_u} Q_u \tag{20}$$

and $\bar{Q}_\Upsilon = (\iota_T \otimes I_N)' Q_\Upsilon (\iota_T \otimes I_N)$. Let

$$\Upsilon = \frac{y' Q_\Upsilon y}{(NT)^{1/2}} = (NT)^{1/2}(s_p^2 - s_v^2 + (1-\hat{\pi}_p)s_u^2) \tag{21}$$

Then under the null hypothesis that all regression functions are identical $\Upsilon / s_\Upsilon \xrightarrow{D} N(0, 1)$ where

$$\begin{aligned} s_\Upsilon^2 = {} & (\hat{\eta}_\varepsilon - 3s_\varepsilon^4)\mathrm{tr} Q_\Upsilon \odot Q_\Upsilon / NT + s_\varepsilon^4 2\mathrm{tr} Q_\Upsilon Q_\Upsilon / NT \\ & + (\hat{\eta}_u - 3s_u^4)\mathrm{tr}\bar{Q}_\Upsilon \odot \bar{Q}_\Upsilon / NT + s_u^4 2\mathrm{tr}\bar{Q}_\Upsilon \bar{Q}_\Upsilon / NT \\ & + s_\varepsilon^2 s_u^2 4\mathrm{tr}(\iota_T \otimes I_N)' Q_\Upsilon Q_\Upsilon (\iota_T \otimes I_N)/NT \quad \blacksquare \end{aligned} \tag{22}$$

Before we apply these procedures to our cost data, we consider two polar cases for the data generating mechanism of the x's.

Example 1: for each i, x_{it} are perfectly correlated over time. In this case, $P_w = I_{NT}$, $Q_u = (1_T - I_T) \otimes D'D$ and $\hat{\pi}_u \doteq 1$. The matrix P_p interleaves the data so that the first T observations correspond to the first firm, the next T to the second firm and so on. Thus with first-order differencing, $\hat{\pi}_p = 1/T$ and hence $s_p^2 \to \sigma_\varepsilon^2 + (1/T)\sigma_u^2$ since differencing pooled data removes individual effects for all observations except those which are preceded by an observation corresponding to a different firm. In the empirical application in this paper, firm size is highly (though not perfectly) correlated over time so that individual electrical utilities have their observations clustered in the pooled data-set.

Example 2: x_{it} are independent over time. Reorderings within each period and in the pooled data are random. With first-order differencing, $\hat{\pi}_u \to d_0^2 = \frac{1}{2}$ since the firm that is 'closest' to firm i in one period is unlikely to be 'closest' in another period. Furthermore $\hat{\pi}_p \to 1$ since individual effects will rarely be differenced out in the pooled data.

4.3 Empirical Results

We begin by obtaining estimates of the variance components σ_u^2 and σ_ε^2. Using equation (12) (or equivalently (7a) applied equation by equation), we estimate β, where the Z matrix contains the explanatory variables corresponding to the Loglinear Model. Replacing y with $y - Z\hat{\beta}$ in equations (16) and (17) we obtain $s_v^2 = 0{\cdot}0170$, $s_u^2 = 0{\cdot}0145$ and by subtraction $s_\varepsilon^2 = 0{\cdot}0025$. Thus, about 85% of the variance of the residual is attributable to the individual specific effect. To test constancy of parametric effects over time, we insert our estimate of the covariance matrix (13) into the conventional asymptotic chi-square statistic for testing linear restrictions. Our Loglinear Model involves seven regression parameters for each year (see e.g. Table I(a)). Thus a test of equality of regression coefficients over time should be approximately χ_{14}^2 under the null. We obtain a value of 63·18, indicating rejection. Casual comparison of Loglinear Model estimates contained in Tables I(a)–(c) would suggest that they are not too different. However, since the residuals are dominated by a firm-specific effect and the explanatory variables are highly

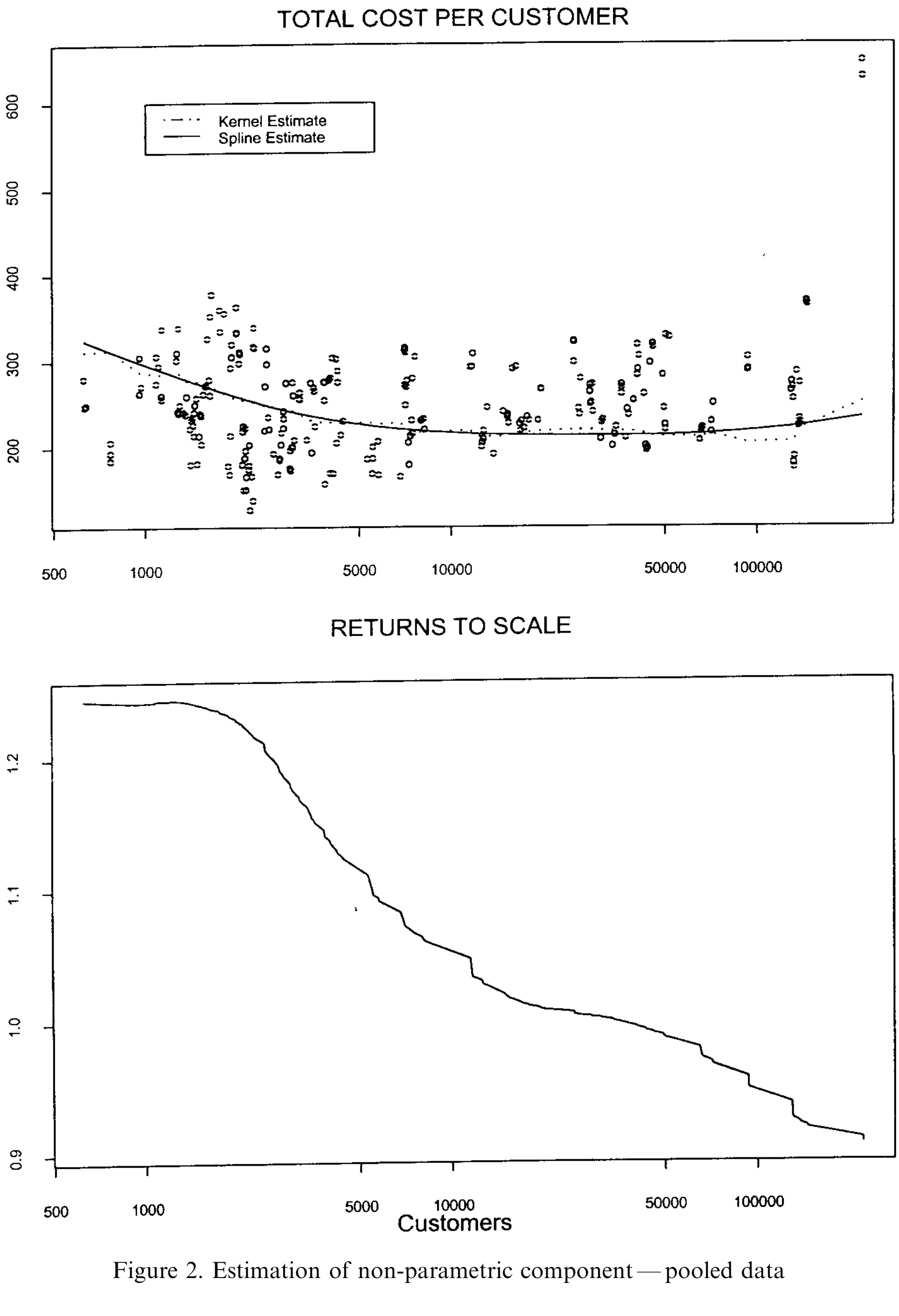

Figure 2. Estimation of non-parametric component — pooled data

correlated over time, coefficient estimates are also highly correlated over time. As a result, even small differences are statistically significant.

Next, we apply the test of equality of non-parametric regression functions in Proposition 3. The standardized statistic has a standard normal distribution under the null hypothesis. Our value is 0·139, indicating that the null cannot be rejected. Figure 2 provides kernel and spline estimates of the non-parametric scale effect using all three years of data where the estimated parametric effects have been removed using the Loglinear Model. It also illustrates the estimated scale economies as a function of the number of customers. Evidently minimum efficient scale is achieved by firms with approximately 20,000 customers. Unit costs appear flat or increase slightly for larger firms with the exception of the largest distributor, which has much higher unit costs.

5. CONCLUSIONS

A central objective of this paper has been to estimate scale economies of electricity distribution under relatively weak functional form assumptions. We have done this by implementing variants of the translog cost function where output — which in our case is the number of customers served — enters non-parametrically, while other variables are parametric. Our tests do not reject homotheticity or linearity in the logs of factor prices. Formal testing rejects the parametric translog in favour of its semiparametric counterpart (equation (1)). Our estimates indicate that minimum efficient scale in Ontario is achieved by utilities with about 20,000 customers. Those utilities which also participated in the delivery of other municipal services had costs that were 7–10% lower, suggesting the presence of economies of scope.

It may be useful to compare our findings to those of other studies. For comparison purposes, we reiterate that our utilities range in size from about 600 to 220,000 customers while sales range from 14 gwh to over 9000 gwh. Giles and Wyatt (1993) analyse data on 60 distributors in New Zealand ranging in size from less than 2000 to over 200,000 customers (Auckland) — a range that is quite comparable to that under study in this paper. Output ranges from about 17 gwh to about 3400 gwh. Giles and Wyatt (1993, p. 378) state that '... any output in the range 500–3500 gwh is essentially consistent with minimum AC'. We have subsequently accessed New Zealand distributor data and found that the implied minimum efficient scale corresponds to utilities with about 30,000 customers.

In their analysis of 100 Norwegian distributors, Salvanes and Tjotta (1994, p. 35) find that '... optimal size comprises plants serving about 20,000 customers and is relatively independent of the level of gwh produced'. Indeed, they also conclude that larger firms exhibit modest decreasing returns to scale. The firms under analysis in the Norwegian study range in size from 655 to 290,560 customers while output ranges from about 11 gwh to 7500 gwh. (See also Salvanes and Tjotta, 1998.) Thus, both the Norwegian and New Zealand results are quite consistent with ours.

Filippini (1996, 1997) analyses 39 Swiss distributors and finds increasing returns to scale throughout his sample. While customer data is apparently not contained in his studies, Filippini defines small utilities to be those with output of about 73 gwh and large utilities to have output of about 300 gwh. Thus, Filippini's 'large utilities' are smaller than those which achieve minimum efficient scale in the Giles and Wyatt study. Comparisons with Salvanes and Tjotta and the current study are somewhat more tenuous because per capita electricity consumption is substantially higher in Norway and Canada. Nevertheless, it would appear that the 'large utilities' Filippini studies are substantially smaller than the large utilities in the Norwegian and

Ontario data.[6] Thus, Filippini's findings of increasing returns to scale throughout his sample may not be inconsistent with the other studies.

The results of our study suggest that *horizontal* mergers between distributors are not likely to produce substantial scale economies *in the operation of their usual wires business*. There are likely to be substantial economies in power procurement, a function which has not been previously performed by most Ontario distributors because the preponderance of electricity has been supplied on an 'as required' basis by Ontario Hydro, the main generator. In Ontario, current restructuring initiatives separate the wires business — which is a natural monopoly and would be regulated — from electricity supply, which would be deregulated. On the other hand, since regulation of the wires business will continue, the presence of a number of distributors within one jurisdiction would help to mitigate the informational asymmetries which encumber the regulator. For example, a larger comparison group would improve the regulator's ability to use techniques such as frontier production function estimation and data envelopment analysis to estimate best practices.[7]

APPENDIX 1

Properties of Permutation and Differencing Matrices

Permutation matrices P have exactly one '1' in every row and column and zeros elsewhere. They are closed under matrix multiplication and $P^{-1} = P'$. Furthermore, tr$P$, trPP, tr$P \odot P$ are all $\leqslant$dim P. For an arbitrary matrix B, $P'BP$ shuffles but otherwise does not alter the diagonal elements of B. (For notation, see last paragraph of Section 2. For more on permutation and related matrices, see Magnus, 1988.)

We may define a more general class of matrices as those that have *at most* one '1' in every row and column and 0's elsewhere. Such matrices are also closed under multiplication and G' is the pseudo-inverse where G is any general permutation matrix; tr$G$, trGG, tr$G \odot G$ are all $\leqslant$dim G. Suppose G is of dimension $NT \times NT$, and define $\bar{G} = (\iota'_T \otimes I_N)G(\iota_T \otimes I_N)$. Then sup $\bar{G} \leqslant T$, tr$\bar{G} \leqslant TN$, tr$\bar{G} \odot \bar{G} \leqslant T^2N$, tr$\bar{G}\bar{G} \leqslant T^2N$.

Within the set of general permutation matrices are the lag matrices. Suppose $i > 0$ and define L'_i to have 0's everywhere except on the ith diagonal above the main diagonal where it has 1's. If $i < 0$, L'_i has 1's on the ith diagonal below the main diagonal. L_0 is defined to be the usual identity matrix, $L'_i = L_{-i}$ and $L_i L_j \doteq L_{i+j}$.

Optimal differencing weights have the property $\Sigma_j d_j d_{j+k} = -1/2m, k = 1, \ldots, m$ (see Hall *et al.*, 1990). Given the order of differencing m, the differencing matrix D may be written as $D = d_0 L_0 + d_1 L'_1 + \cdots + d_m L'_m$. Thus, D is a finite linear combination of general permutation matrices as are $D'D$, DD', $D'DD'D$, DG and $D'G$ where G is any general permutation matrix. The matrix $D'D$ has a symmetric band structure with (except for end effects), ones on the main diagonal, $-1/2m$ on the m adjacent diagonals and zeros elsewhere. That is, $D'D \doteq L_0 - (1/2m)(L_1 + L'_1 + \cdots L_m + L'_m)$ so that tr$(D'D) \doteq N$. The matrix $D'DD'D$ has a symmetric band

[6] Per capita annual electricity consumption in Switzerland and New Zealand are quite comparable at 7346 and 8865 kwh per year respectively. For Norway and Canada the figures are 24,033 and 16,413 kwh. (Source: International Energy Agency, http://www.iea.org/stat.htm, 1996 data.)

[7] Green and Jackson (1994) use frontier production function techniques to analyse distributor data in the United Kingdom, but their data-set consists of only 12 observations.

structure with $1 + 1/2m$ on the main diagonal so that $\text{tr}(D'DD'D) \doteq N(1 + 1/2m)$. The first m diagonals adjacent to the main diagonal take the value:

$$-\frac{1}{m} + j\frac{1}{4m^2}$$

$j = 2m - 2, 2m - 3, \ldots, m - 1$. The next m diagonals take the value: $j(1/4m^2), j = m, m - 1, \ldots, 1$. The remainder of the matrix is 0.

Throughout the paper, traces divided by N converge in probability by virtue of our assumptions on the data-generating mechanisms for the x's. (In Section 3, Single Equation Analysis x_{it} is independent of x_{js} if $i \neq j$ or $s \neq t$. In Section 4, Panel Data Analysis, $(x_{i1}, \ldots, x_{iT})$ is independent of $(x_{j1}, \ldots, x_{jT})$ if $i \neq j$.)

Asymptotic Distribution of Differencing Estimator

In the following, brief justification is provided for equation (7a). (For expositional purposes, the 't' subscript is dropped.) See also Yatchew (1997, 1998a, pp. 670–2, 694–9).

Using equation 6 note that since differencing removes the non-parametric effect f we have $Dy \cong DZ\beta + Dv$. Next, write $Z = g(X) + U$ where g is a smooth vector function of conditional means of each *parametric* explanatory variable given the *non-parametric* variable; $g(X)$ is an $N \times k$ matrix whose ith row contains the components of g evaluated at x_i. Since differencing removes g, $DZ \cong DU$. Note that $U'U/n \xrightarrow{P} \Sigma_{z|x}$. Now

$$\hat{\beta} = [(DZ)'DZ]^{-1}(DZ)'Dy \cong \beta + [(DU)'DU]^{-1}(DU)'Dv$$

hence

$$\text{Cov}(n^{1/2}(\hat{\beta} - \beta)) \cong \sigma_v^2 \left(\frac{U'D'DU}{n}\right)^{-1} \left(\frac{U'D'DD'DU}{n}\right) \left(\frac{U'D'DU}{n}\right)^{-1}$$

Since $D'D$ has (except for end effects) ones on the main diagonal, $U'D'DU/n \xrightarrow{P} \Sigma_{z|x}$. Furthermore, since $D'DD'D$ has (except for end effects) $1 + 1/2m$ on the main diagonal $U'D'DD'DU/n \xrightarrow{P} (1 + 1/2m)\Sigma_{z|x}$. Thus

$$\text{Cov}(n^{1/2}(\hat{\beta} - \beta)) \xrightarrow{P} \sigma_v^2 \left(1 + \frac{1}{2m}\right)\Sigma_{z|x}^{-1}$$

Lemma 1: Define the NT-dimensional stacked vector v as in equation (10) with $\text{Cov}(v) = \sigma_\varepsilon^2 I_{NT} + \sigma_u^2(1_T \otimes I_N)$. Let Q be an $NT \times NT$ symmetric matrix, $\bar{Q} = (\iota_T \otimes I_N)'Q(\iota_T \otimes I_N)$. Then $E(v'Qv) = \sigma_\varepsilon^2 \text{tr}Q + \sigma_u^2 \text{tr}\bar{Q}$ and

$$\begin{aligned} \text{Var}(v'Qv) = {} & (\eta_\varepsilon - 3\sigma_\varepsilon^4)\text{tr}Q \odot Q + \sigma_\varepsilon^4 2\text{tr}QQ \\ & + (\eta_u - 3\sigma_u^4)\text{tr}\bar{Q} \odot \bar{Q} + \sigma_u^4 2\text{tr}\bar{Q}\bar{Q} \\ & + \sigma_\varepsilon^2\sigma_u^2 4\text{tr}(\iota_T \otimes I_N)'QQ(\iota_T \otimes I_N) \quad \blacksquare \end{aligned}$$

Comment on Lemma 1: Lemma 1 can be used to calculate asymptotic variances of the quadratic forms in s_v^2, s_u^2 and s_p^2 as well as of test statistic Υ in Proposition 3. To prove Lemma 1 we will use the following. Suppose $\vartheta = (\vartheta_1, \ldots, \vartheta_\xi)'$ where $E\vartheta_i = 0$, $\mathrm{Var}(\vartheta_i) = \sigma_\vartheta^2$, $E\vartheta_i^4 = \eta_\vartheta$ and ϑ has covariance matrix $\sigma_\vartheta^2 I_\xi$. If A is a symmetric matrix, then $E(\vartheta' A \vartheta) = \sigma_\vartheta^2 \mathrm{tr} A$ and $\mathrm{Var}(\vartheta' A \vartheta) = (\eta_\vartheta - 3\sigma_\vartheta^4)\mathrm{tr} A \odot A + \sigma_\vartheta^4 2\mathrm{tr} AA$. For results of this type see e.g. Schott (1997) or they may be proved directly. ■

Proof of Lemma 1: Rewrite $v'Qv = \varepsilon' Q \varepsilon + u'Qu + 2\varepsilon' Qu = \varepsilon' Q \varepsilon + u_0' \bar{Q} u_0 + 2\varepsilon' Qu$ where $u = \iota_T \otimes u_0$ and $u_0 = (u_1, \ldots, u_N)'$ contains the individual effects with which individuals are endowed at birth. Note that the three terms in the expansion are mutually uncorrelated. Using the Comment above obtain $E(\varepsilon' Q \varepsilon)$ and $E(u_0' \bar{Q} u_0)$. Collect terms to obtain $E(v'Qv)$. Again referring to the Comment above, calculate $\mathrm{Var}(\varepsilon' Q \varepsilon)$ and $\mathrm{Var}(u_0' \bar{Q} u_0)$. Note that $\mathrm{Var}(\varepsilon' Qu) = E(u' Q \varepsilon \varepsilon' Qu) = \sigma_\varepsilon^2 E_u(u' QQu) = \sigma_\varepsilon^2 \sigma_u^2 \mathrm{tr}(\iota_T' \otimes I_N) QQ(\iota_T \otimes I_N)$. Collect terms to obtain $\mathrm{Var}(v'Qv)$. ■

Lemma 2: Suppose (x_i, ε_i), $i = 1, \ldots, \xi$ are i.i.d. The x_i have density bounded away from zero on the unit interval and $\varepsilon_i | x_i \sim (0, \sigma_\varepsilon^2)$. Assume data have been reordered so that $x_1 \leqslant \cdots \leqslant x_\xi$. Define $\tilde{f} = (f(x_1), \ldots, f(x_\xi))'$ where the function f has a bounded first derivative. Let D be a differencing matrix of, say, order m. Then $\tilde{f}' D' D \tilde{f} = O_P(\xi^{-1+\delta})$ and $\mathrm{Var}(\tilde{f}' D' D \varepsilon) = O_P(\xi^{-1+\delta})$ where δ is positive and arbitrarily close to 0. ■

Proof of Lemma 2: The results follow immediately from Yatchew (1997, Appendix, equations (A.2 and (A.3)). ■

Lemma 3: Let G_N be a sequence of $N \times N$ general permutation matrices such that $\mathrm{tr} G_N / N$ and $\mathrm{tr} G_N G_N / N$ converge in probability to constants λ and γ respectively. Let $\vartheta_i \sim (0, \sigma_\vartheta^2)$ be i.i.d. random variables with finite fourth-moment η_ϑ and define $\vartheta = (\vartheta_1, \ldots, \vartheta_N)'$. Then

$$N^{1/2}\left(\frac{1}{N}\vartheta' G_N \vartheta - \frac{\sigma_\vartheta^2}{N}\mathrm{tr} G_N\right) \overset{D}{\rightarrow} N(0, \lambda(\eta_\vartheta - 3\sigma_\vartheta^4) + \gamma 2 \sigma_\vartheta^4)$$

Proof of Lemma 3: Rewrite $G_N = \Lambda_N + \Gamma_N$ where Λ_N is a diagonal matrix (with 1's and 0's on the diagonal) and Γ_N which has 0's on the main diagonal. Note that $\vartheta' \Lambda_N \vartheta$ and $\vartheta' \Gamma_N \vartheta$ are uncorrelated. Since $\mathrm{tr}\Lambda_N / N = \mathrm{tr} G_N / N \overset{P}{\rightarrow} \lambda$ we have

$$N^{1/2}\left(\frac{1}{N}\vartheta' \Lambda_N \vartheta - \frac{\sigma_\vartheta^2}{N}\mathrm{tr} G_N\right) \overset{D}{\rightarrow} N(0, \lambda(\eta_\vartheta - \sigma_\vartheta^4))$$

Next, note that the eigenvalues of Γ_N are bounded in absolute value by 1 and that each row of Γ_N contains at most one non-zero element (which equals 1). Further,

$$\mathrm{tr}\Gamma_N \Gamma_N / N \overset{P}{\rightarrow} \gamma - \lambda$$

We may now apply de Jong (1987, Theorem 5.2) to conclude that

$$N^{1/2}\frac{\vartheta' \Gamma_N \vartheta}{N} \overset{D}{\rightarrow} N(0, 2(\gamma - \lambda)\sigma_\vartheta^4)$$

and the result of Lemma 3 follows immediately. ■

Comments on Lemma 3: Suppose P is a permutation matrix. Then

$$P'D'DP \doteq L_0 - \frac{1}{2m}[P'L_1P + P'L_1'P + \cdots + P'L_mP + P'L_m'P]$$

Note that $P'L_iP$ and $PL_i'P$ are matrices of the form used in the quadratic form of Lemma 3 since each is a (general) permutation matrix. More generally, let P_A, P_B be (general) permutation matrices and consider:

$$P_A'D'DP_B \doteq P_A'L_0P_B - \frac{1}{2m}[P_A'L_1P_B + P_A'L_1'P_B + \cdots + P_A'L_mP_B + P_A'L_m'P_B]$$

which is a weighted combination of matrices that satisfy the form used in the quadratic form of Lemma 3 since $P_A'L_iP_B$ and $P_A'L_i'P_B$ are (general) permutation matrices.

Similarly, by straightforward expansion and regrouping of terms, it can be shown that $P_A'D'P_AP_B'DP_B$ and $P_B'D'P_BP_A'DP_A$ can be rewritten as a weighted sum of matrices of the form used in the quadratic form of Lemma 3. ■

Proof of Proposition 1: In the following we make use of the above 'Properties of Permutations and Differencing Matrices'.

Consistency of s_v^2: $\mathrm{tr}Q_v \doteq NT$ and $\mathrm{tr}\bar{Q}_v \doteq NT$ where $\bar{Q}_v = (\imath_T' \otimes I_N)Q_v(\imath_T \otimes I_N)$. Note that $\mathrm{tr}Q_v \odot Q_v$, $\mathrm{tr}Q_vQ_v$, $\mathrm{tr}\bar{Q}_v \odot \bar{Q}_v$, $\mathrm{tr}\bar{Q}_v\bar{Q}_v$ and $\mathrm{tr}(\imath_T \otimes I_N)'Q_vQ_v(\imath_T \otimes I_N)$ are $O(N)$. Apply Lemma 2 to conclude that $s_v^2 - v'Q_vv/NT \xrightarrow{P} 0$. Apply Lemma 1 to conclude that $E(v'Q_vv)/NT \doteq \sigma_v^2$ and that $\mathrm{Var}(v'Q_vv) = O(N)$. Hence $\mathrm{Var}(s_v^2) \to 0$ and the estimator is consistent.

Consistency of s_p^2: $\mathrm{tr}Q_p \doteq NT$ and $\mathrm{tr}\bar{Q}_p = NT\hat{\pi}_p$ where $\bar{Q}_p = (\imath_T' \otimes I_N)Q_p(\imath_T \otimes I_N)$ so that conditional on the x's, $E(v'Q_pv/NT) \doteq \sigma_\varepsilon^2 + \hat{\pi}_p\sigma_u^2$. Note that $\mathrm{tr}Q_p \odot Q_p$, $\mathrm{tr}Q_pQ_p$, $\mathrm{tr}\bar{Q}_p \odot \bar{Q}_p$, $\mathrm{tr}\bar{Q}_p\bar{Q}_p$ and $\mathrm{tr}(\imath_T \otimes I_N)'Q_pQ_p(\imath_T \otimes I_N)$ are $O(N)$. Now follow the above proof of consistency of s_v^2.

Consistency of s_u^2: Q_u has diagonal elements 0, thus $\mathrm{tr}Q_u = \mathrm{tr}Q_u \odot Q_u = 0$. Write $\bar{Q}_u = (\imath_T' \otimes I_N)Q_u(\imath_T \otimes I_N) = \Sigma_{s\neq t}P_s'D'P_sP_t'DP_t$ where $P_1, \ldots, P_T$ are the $N \times N$ permutation matrices that make up the diagonal blocks of P_w. We have $\mathrm{tr}\bar{Q}_u = NT(T-1)\hat{\pi}_u$ and conditional on the x's $E(v'\bar{Q}_uv/(NT(T-1)\hat{\pi}_u)) \doteq \sigma_u^2$. Next, note that $\mathrm{tr}Q_uQ_u = \Sigma_{s\neq t}\mathrm{tr}P_s'DD'P_sP_t'DD'P_t$ $\mathrm{tr}\bar{Q}_u \odot \bar{Q}_u$, $\mathrm{tr}\bar{Q}_u\bar{Q}_u$ and $\mathrm{tr}(\imath_T' \otimes I_N)Q_uQ_u(\imath_T \otimes I_N)$ are $O(N)$. Now follow the above proof of consistency s_v^2. ■

Proof of Proposition 2: Using arguments similar to Lemma 2 it can be shown that differencing removes the nonparametric effect quickly enough so that

$$\frac{1}{NT}\sum_t (P_t'DP_ty_t \odot P_t'DP_ty_t)'(P_t'DP_ty_t \odot P_t'DP_ty_t) - (P_t'DP_tv_t \odot P_t'DP_tv_t)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \xrightarrow{P} 0$$

$$\frac{1}{NT(T-1)}\sum_{s\neq t} (P_s'DP_sy_s \odot P_s'DP_sy_s)'(P_t'DP_ty_t \odot P_t'DP_ty_t) - (P_s'DP_sv_s \odot P_s'DP_sv_s)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \xrightarrow{P} 0$$

We will show that

$$\frac{1}{NT}\sum_{t}(P_t'DP_tv_t \odot P_t'DP_tv_t)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \xrightarrow{P} \frac{1}{2}\eta_e + \frac{1}{2}\eta_\epsilon + 6\sigma_u^2\sigma_\varepsilon^2 + \frac{3}{2}\sigma_u^4 + \frac{3}{2}\sigma_\varepsilon^4$$
$$\frac{1}{NT(T-1)}\sum_{s\neq t}(P_s'DP_sv_s \odot P_s'DP_sv_s)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \xrightarrow{P} \frac{1}{2}\pi_u\eta_u + 2\sigma_u^2\sigma_\varepsilon^2 + \frac{3}{2}\pi_u\sigma_u^4 + \sigma_\varepsilon^4 \quad \text{(A1.1)}$$

Existence of fourth-order moments of u_i and ε_{it} is sufficient to ensure that the LHS of the above equations converge to constants. To determine those constants, we need to compute certain expectations. Rewrite $P_t'DP_tv_t = P_t'(d_0L_0 + d_1L_1')P_tv_t = (v_t + A_tv_t)/\sqrt{2}$ noting that $A_t \equiv P_t'L_1'P_t$ is a (general) permutation matrix with the important property that it has 0's on the diagonal. Note also that $A_tv_t \odot A_tv_t = A_t(v_t \odot v_t)$ and $(v_t \odot A_tv_t)'(v_t \odot A_tv_t) = (v_t \odot v_t)'(A_tv_t \odot A_tv_t) = (v_t \odot v_t)'A_t(v_t \odot v_t)$.

Define $\eta_v = Ev_{it}^4 = \eta_u + 6\sigma_u^2\sigma_\varepsilon^2 + \eta_\varepsilon$ and note that $(Ev_{it}^2)^2 = (\sigma_u^2 + \sigma_\varepsilon^2)^2 = \sigma_u^4 + 2\sigma_\varepsilon^2\sigma_u^2 + \sigma_\varepsilon^4$. Consider

$$\begin{aligned} &E(P_t'DP_tv_t \odot P_t'DP_tv_t)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \\ &= \tfrac{1}{4}E(v_t \odot v_t + 2v_t \odot A_tv_t + A_t(v_t \odot v_t))'(v_t \odot v_t + 2v_t \odot A_tv_t + A_t(v_t \odot v_t)) \end{aligned} \quad \text{(A1.2)}$$

which can be expanded and the expectations evaluated term by term. The non-zero terms are:

(i) $E(v_t \odot v_t)'(v_t \odot v_t) = NEv_{it}^4$
(ii) $E(v_t \odot v_t)'A_t(v_t \odot v_t) \doteq N(Ev_{it}^2)^2$
(iii) $E4(v_t \odot A_tv_t)'(v_t \odot A_tv_t) = E4(v_t \odot v_t)'(A_tv_t \odot A_tv_t) \doteq N4(Ev_{it}^2)^2$
(iv) $E(v_t \odot v_t)'A_t'A_t(v_t \odot v_t) \doteq E(v_t \odot v_t)'(v_t \odot v_t) = NEv_{it}^4$

Now collect terms to conclude that, except for 'end effects', equation (A1.2) equals

$$\tfrac{1}{4}N[2(\eta_u + 6\sigma_u^2\sigma_\varepsilon^2 + \eta_\varepsilon) + 6(\sigma_u^4 + 2\sigma_u^2\sigma_\varepsilon^2 + \sigma_\varepsilon^4)] = N[\tfrac{1}{2}\eta_u + \tfrac{1}{2}\eta_\varepsilon + 6\sigma_u^2\sigma_\varepsilon^2 + \tfrac{3}{2}\sigma_u^4 + \tfrac{3}{2}\sigma_\varepsilon^4]$$

Next, for $s \neq t$ consider

$$\begin{aligned} &E(P_s'DP_sv_s \odot P_s'DP_sv_s)'(P_t'DP_tv_t \odot P_t'DP_tv_t) \\ &= \tfrac{1}{4}E(v_s \odot v_s + 2v_s \odot A_sv_s + A_s(v_s \odot v_s))'(v_t \odot v_t + 2v_t \odot A_tv_t + A_t(v_t \odot v_t)) \end{aligned} \quad \text{(A1.3)}$$

which can be expanded and expectations evaluated term by term. The non-zero terms are:

(i) $E(v_s \odot v_s)'(v_t \odot v_t) = E(u_0 \odot u_0 + 2u_0 \odot \varepsilon_s + \varepsilon_s \odot \varepsilon_s)'(u_0 \odot u_0 + 2u_0 \odot \varepsilon_t + \varepsilon_t \odot \varepsilon_t)$
$= N(\eta_u + 2\sigma_u^2\sigma_\varepsilon^2 + \sigma_\varepsilon^4)$
(ii) $E(v_s \odot v_s)'A_t(v_t \odot v_t) \doteq N(\sigma_u^4 + 2\sigma_\varepsilon^2\sigma_u^2 + \sigma_\varepsilon^4)$
(iii) $E4(v_s \odot A_sv_s)'(v_t \odot A_tv_t) = 4E(u_0 \odot A_su_0)'(u_0 \odot A_tu_0) = 4E(u_0 \odot u_0)'(A_su_0 \odot A_tu_0)$
$= 4\sigma_u^4\text{tr}A_s'A_t$
(iv) $E(v_s \odot v_s)'A_s'(v_t \odot v_t) \doteq N(\sigma_u^4 + 2\sigma_\varepsilon^2\sigma_u^2 + \sigma_\varepsilon^4)$
(v) $E(v_s \odot v_s)'A_s'A_t(v_t \odot v_t) = E(u_0 \odot u_0 + 2u_0 \odot \varepsilon_s + \varepsilon_s \odot \varepsilon_s)'A_s'A_t(u_0 \odot u_0 + 2u_0 \odot \varepsilon_t + \varepsilon_t \odot \varepsilon_t)$
$\doteq N(\eta_u\text{tr}A_s'A_t/N + \sigma_u^4(1 - \text{tr}A_s'A_t/N) + 2\sigma_u^2\sigma_\varepsilon^2 + \sigma_\varepsilon^4)$

Now collect terms to conclude that, except for end effects, equation (A1.3) equals

$$\frac{1}{4}N\left[\eta_u\left(1+\frac{\mathrm{tr}A_s'A_t}{N}\right)+3\sigma_u^4\left(1+\frac{\mathrm{tr}A_s'A_t}{N}\right)+8\sigma_u^2\sigma_\varepsilon^2+4\sigma_\varepsilon^4\right]$$

Note that $1+\mathrm{tr}A_s'A_t/N=2\mathrm{tr}P_s'D'P_sP_t'DP_t/N$ and that $\hat{\pi}_u=\mathrm{tr}\Sigma_{s\neq t}P_s'D'P_sP_t'DP_t/NT(T-1)$. Thus we have shown that equation (A1.1) holds.

Using equation (A1.1) we may now conclude that

$$\eta_u=\left(\mathrm{plim}\left(\frac{2}{NT(T-1)}\sum_{s\neq t}(P_s'DP_sy_s\odot P_s'DP_sy_s)'(P_t'DP_ty_t\odot P_t'DP_ty_t)\right)-4\sigma_u^2\sigma_\varepsilon^2-3\pi_u\sigma_u^4-2\sigma_\varepsilon^4\right)\Big/\pi_u$$

$$\eta_\varepsilon=\mathrm{plim}\left(\frac{2}{NT}\sum_t(P_t'DP_ty_t\odot P_t'DP_ty_t)'(P_t'DP_ty_t\odot P_t'DP_ty_t)\right)-\eta_u-12\sigma_u^2\sigma_\varepsilon^2-3\sigma_u^4-3\sigma_\varepsilon^4$$

Replacing quantities on the right-hand sides with consistent estimates yields consistent estimators of η_u and η_ε. ■

Proof of Proposition 3: Using Lemma 2, conclude that $(NT)^{1/2}(s_v^2-v'Q_vv/NT)\xrightarrow{P}0$, $(NT)^{1/2}(s_p^2-v'Q_pv/NT)\xrightarrow{P}0$, $(NT)^{1/2}(s_u^2-v'Q_uv/(\pi_uNT(T-1)))\xrightarrow{P}0$ in which case using equation (21) we may conclude that $(\Upsilon-v'Q_\Upsilon v/(NT)^{1/2})\xrightarrow{P}0$.

Rewrite $v'Q_vv$, $v'Q_pv$ and $v'Q_uv$ as finite weighted combinations of terms of the form $\varepsilon_t'G\varepsilon_t$, $u_0'Gu_0$, $\varepsilon_s'G\varepsilon_t$ and $u_0'G\varepsilon_t$ where G is a general permutation matrix of dimension $N\times N$ (see comments following Lemma 3 above). The number of such terms depends on T and m but not on N. Now apply Lemma 3 to terms of the first and second type. For terms of the form $\varepsilon_s'G\varepsilon_t$ and $u_0'G\varepsilon_t$ note that since G has at most one '1' in each row and column, such terms are sums of independent random variables so that a conventional CLT may be applied. Using equation (20) conclude that $v'Q_\Upsilon v/(NT)^{1/2}$ and hence $y'Q_\Upsilon y/(NT)^{1/2}$ are asymptotically normal. Recall $\mathrm{tr}Q_v\doteq NT$, $\mathrm{tr}Q_p\doteq NT$ and Q_u has diagonal elements 0. Hence, conditional on the x's, $E(v'Q_\Upsilon v/(NT)^{1/2})=\mathrm{tr}Q_\Upsilon/(NT)^{1/2}\doteq 0$. Using Lemma 1, $\mathrm{Var}(v'Q_\Upsilon v/(NT)^{1/2})$ may be calculated. Replacing various moments with consistent estimates yields equation (22). ■

APPENDIX 2. VARIABLE DEFINITIONS AND SUMMARY STATISTICS

		1993				1994				1995			
		Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Raw data													
TC	total costs ($mill/annum)	6·73	17·22	0·137	137·0	7·02	17·44	0·144	138·0	7·02	17·06	0·153	134·0
CUST	number of customers	23538	41012	629	219376	23718	41236	624	219723	23895	41474	636	220215
WAGE	wage of lineman ($/hour)	21·35	1·92	16·50	24·53	21·54	1·88	16·75	24·53	21·64	1·86	17·00	24·53
TOTPLANT	accum. gross investment ($mill.)	53·08	116·0	0·891	767·0	55·80	121·0	0·945	776·0	59·98	133·0	0·970	881·0
PUC	public utility commission dummy	0·457	0·501	0	1	0·457	0·501	0	1	0·457	0·501	0	1
KWH	kilowatt hour sales (gwh/yr)	791	1630	14·2	9220	799	1640	14·51	9310	802	1640	15·3	9150
LIFE	fixed assets remaining lifetime (yrs)	14·41	2·303	8·90	21·2	13·41	2·303	7·90	20·2	12·41	2·303	6·90	19·2
LF	load factor (%)	70·94	5·235	37·59	82·54	70·94	5·235	37·59	82·54	70·94	5·235	37·59	82·54
KMWIRE	kilometres of distribution wire	311·24	529·83	10·0	3000·0	311·79	529·75	10·2	3000·1	312·22	529·73	10·8	3000·2
Dependent variable: y													
tc	log(*TC*/*CUST*)	5·439	0·236	4·816	6·439	5·474	0·243	4·894	6·439	5·500	0·216	5·080	6·410
Non-parametric independent variable: x													
cust	log(*CUST*)	8·875	1·540	6·444	12·299	8·886	1·540	6·436	12·300	8·894	1·539	6·455	12·300
Parametric independent variables													
PUC		0·457	0·501	0	1	0·457	0·501	0	1	0·457	0·501	0	1
wage	log(*WAGE*)	3·057	0·092	2·803	3·20	3·066	0·089	2·818	3·20	3·070	0·088	2·833	3·20
pcap	log(*TOTPLANT*/*KMWIRE*)	11·686	0·454	10·305	12·63	11·732	0·454	10·381	12·68	11·792	0·458	10·430	12·76
kwh	log(*KWH*/*CUST*)	10·088	0·285	9·392	10·70	10·085	0·295	9·340	10·72	10·078	0·302	9·313	10·74
life	log(*LIFE*)	2·655	0·162	2·186	3·05	2·581	0·175	2·067	3·01	2·501	0·190	1·932	2·95
lf	log(*LF*)	4·258	0·087	3·627	4·41	4·259	0·087	3·627	4·41	4·258	0·087	3·627	4·41
kmwire	log(*KMWIRE*/*CUST*)	−4·292	0·317	−4·989	−3·626	−4·292	0·312	−4·994	−3·630	−4·292	0·313	−5·000	−3·607

ACKNOWLEDGEMENTS

The author is grateful to John Rust, Kjell Salvanes, Sigve Tjotta and two anonymous referees for their helpful comments.

REFERENCES

Ahn, H. and J. Powell (1993), 'Semiparametric estimation of censored selection models with a non-parametric selection mechanism', *Journal of Econometrics*, **58**, 2–29.

Baltagi, B., J. Hidalgo and Q. Li (1996), 'A nonparametric test for poolability using panel data', *Journal of Econometrics*, **75**, 345–367.

Cox, D. and E. Koh (1989), 'A smoothing spline based test of model adequacy in polynomial regression', *Annals of the Institute of Statistical Mathematics*, **41**, 383–400.

de Jong, P. (1987), 'A central limit theorem for generalized quadratic forms', *Probab. Th. Rel. Fields*, **75**, 261–277.

Filippini, M. (1996), 'Economies of scale and utilization', *Applied Economics*, **28**, 543–550.

Filippini, M. (1997), *Elements of the Swiss Market for Electricity*, Physica Verlag, Heidelberg.

Giles, D. and N. S. Wyatt (1993), 'Economies of scale in the New Zealand electricity distribution industry', in P. C. B. Phillips (ed.), *Models, Methods and Applications of Econometrics*, Blackwell, Oxford, 370–382.

Green, R. and A. Jackson (1994), 'Electricity distribution: a comparative efficiency study of the regional electricity companies', Department of Applied Economics, Cambridge University, Cambridge, UK.

Hall, P., J. W. Kay and D. M. Titterington (1990), 'Asymptotically optimal difference-based estimation of variance in nonparametric regression', *Biometrika*, **77**, 521–528.

Hjalmarsson, L. and A. Veiderpass (1992a), 'Productivity in Swedish electricity retail distribution', *Scandinavian Journal of Economics*, **94** (Supplement), 193–205.

Hjalmarsson, L. and A. Veiderpass (1992b), 'Efficiency and ownership in Swedish electricity retail distribution', *Journal of Productivity Analysis*, **3**, 7–23.

Kwoka, J. (1997), *Power Structure: Ownership, Integration and Competition in the U.S. Electricity Industry*, Kluwer Academic Publishers, Boston, MA.

Magnus, J. R. (1988), *Linear Structures*, Charles Griffin and Co., London/Oxford University Press, New York.

Nelson, R. (1990), 'The effects of competition on publicly-owned firms: evidence from the municipal electric industry in the U.S.', *International Journal of Industrial Organization*, **8**, 37–51.

Neuberg, L. (1977), 'Two issues in the municipal ownership of electric power distribution systems', *Bell Journal of Economics*, **8**, 302–323.

Ontario Hydro (1995), *Accounting for Municipal Electric Utilities in Ontario*.

Pollitt, M. J. (1995), *Ownership and Performance in Electric Utilities: The International Evidence on Privatization and Efficiency*, Oxford University Press, Oxford.

Powell, J. (1987), 'Semiparametric estimation of bivariate latent variable models', Working Paper 8704, Social Systems Research Institute of Wisconsin, University of Wisconsin, Madison.

Rice, J. (1984), 'Bandwidth choice for nonparametric regression', *Annals of Statistics*, **12**, 1215–1230.

Salvanes, K. and S. Tjotta (1994), 'Productivity differences in multiple output industries', *Journal of Productivity Analysis*, **5**, 23–43.

Salvanes, K. and S. Tjotta (1998), 'A test for natural monopoly with application to Norwegian electricity distribution', *Review of Industrial Organization*, **13**, 669–685.

Schott, J. R. (1997), *Matrix Analysis for Statistics*, Wiley, New York.

Yatchew, A. J. (1988), 'Some tests of nonparametric regressions models', in W. Barnett, E. Berndt and H. White (eds), *Dynamic Econometric Modelling*, Proceedings of the Third International Symposium in Economic Theory and Econometrics, Cambridge University Press, Cambridge, 121–135.

Yatchew, A. J. (1997), 'An elementary estimator of the partial linear model', *Economics Letters*, **57**, 135–143, additional tables contained in *Economics Letters* 1998, **59**, 403–405.

Yatchew, A. J. (1998a), 'Nonparametric regression techniques in economics', *Journal of Economic Literature*, **XXXVI**, 669–721.

Yatchew, A. J. (1998b), 'An elementary nonparametric differencing test of equality of regression functions', *Economics Letters*, in press.

