sets tend to be consecutive after pooling and reordering, then the *covariance* between the two terms will be large. In particular, the covariance is approximately $\sigma_\varepsilon^4(1-\pi)$, where π equals the probability that consecutive observations in the pooled reordered data set come from *different* populations.

It follows that under $H_o: f_A = f_B$,

$$\Upsilon \xrightarrow{D} N\left(0, 2\pi\sigma_\varepsilon^4\right). \tag{1.5.6}$$

For example, if reordering the pooled data is equivalent to stacking data sets A and B – because the two sets of x's, x_A and x_B, do not intersect – then $\pi \cong 0$ and indeed the statistic Υ becomes degenerate. This is not surprising, since observing nonparametric functions over different domains cannot provide a basis for testing whether they are the same. If the pooled data involve a simple interleaving of data sets A and B, then $\pi \cong 1$ and $\Upsilon \to N(0, 2\sigma_\varepsilon^4)$. If x_A and x_B are independent of each other but have the same distribution, then for the pooled reordered data the probability that consecutive observations come from different populations is ½ and $\Upsilon \to N(0, \sigma_\varepsilon^4)$.[4] To implement the test, one may obtain a consistent estimate $\hat{\pi}$ by taking the proportion of observations in the pooled reordered data that are preceded by an observation from a different population.

1.6 Empirical Application: Scale Economies in Electricity Distribution[5]

To illustrate these ideas, consider a simple variant of the Cobb–Douglas model for the costs of distributing electricity

$$\begin{aligned} tc = {} & f(cust) + \beta_1 wage + \beta_2 pcap \\ & + \beta_3 PUC + \beta_4 kwh + \beta_5 life + \beta_6 lf + \beta_7\, kmwire + \varepsilon \end{aligned} \tag{1.6.1}$$

where *tc* is the log of total cost per customer, *cust* is the log of the number of customers, *wage* is the log wage rate, *pcap* is the log price of capital, *PUC* is a dummy variable for public utility commissions that deliver additional services and therefore may benefit from economies of scope, *life* is the log of the remaining life of distribution assets, *lf* is the log of the load factor (this measures capacity utilization relative to peak usage), and *kmwire* is the log of kilometers of distribution wire per customer. The data consist of 81 municipal distributors in Ontario, Canada, during 1993. (For more details, see Yatchew, 2000.)

[4] For example, distribute n men and n women randomly along a stretch of beach facing the sunset. Then, for any individual, the probability that the person to the left is of the opposite sex is ½. More generally, if x_A and x_B are independent of each other and have different distributions, then π depends on the relative density of observations from each of the two populations.

[5] Variable definitions for empirical examples are contained in Appendix E.

Because the data have been reordered so that the nonparametric variable *cust* is in increasing order, first differencing (1.6.1) tends to remove the nonparametric effect f. We also divide by $\sqrt{2}$ so that the residuals in the differenced Equation (1.6.2) have the same variance as those in (1.6.1). Thus, we have

$$\begin{aligned}
&[tc_i - tc_{i-1}]/\sqrt{2} \\
&\quad \cong \beta_1[wage_i - wage_{i-1}]/\sqrt{2} + \beta_2[pcap_i - pcap_{i-1}]/\sqrt{2} \\
&\qquad + \beta_3[PUC_i - PUC_{i-1}]/\sqrt{2} + \beta_4[kwh_i - kwh_{i-1}]/\sqrt{2} \\
&\qquad + \beta_5[life_i - life_{i-1}]/\sqrt{2} + \beta_6[lf_i - lf_{i-1}]/\sqrt{2} \\
&\qquad + \beta_7[kmwire_i - kmwire_{i-1}]/\sqrt{2} + [\varepsilon_i - \varepsilon_{i-1}]/\sqrt{2}. \qquad (1.6.2)
\end{aligned}$$

Figure 1.2 summarizes our estimates of the parametric effects β using the differenced equation. It also contains estimates of a pure parametric specification in which the scale effect f is modeled with a quadratic. Applying the specification test (1.4.2), where s^2_{diff} is replaced with (1.3.5), yields a value of 1.50, indicating that the quadratic model may be adequate.

Thus far our results suggest that by differencing we can perform inference on β as if there were no nonparametric component f in the model to begin with. But, having estimated β, we can then proceed to apply a variety of nonparametric techniques to analyze f as if β were known. Such a modular approach simplifies implementation because it permits the use of existing software designed for pure nonparametric models.

More precisely, suppose we assemble the ordered pairs $(y_i - z_i\hat{\beta}_{diff}, x_i)$; then, we have

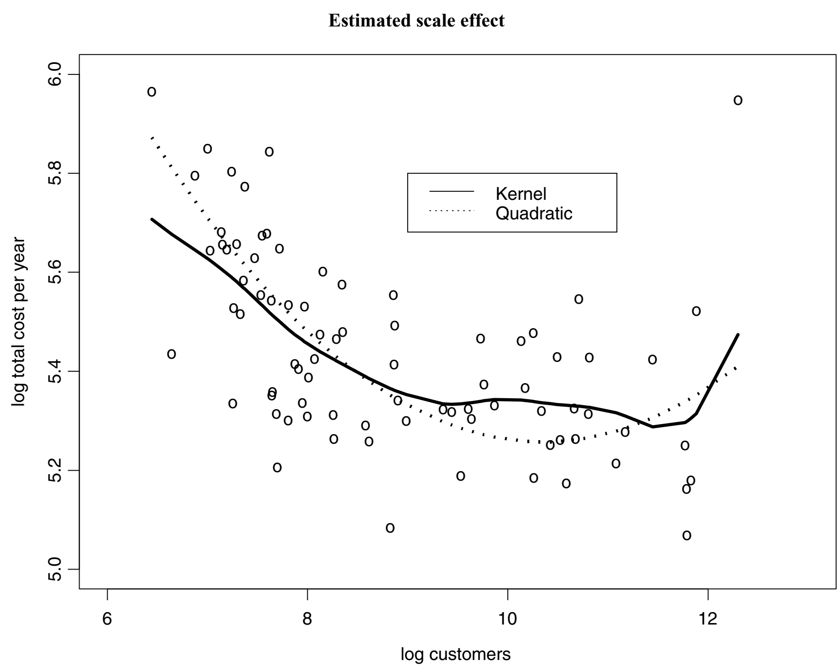
$$y_i - z_i\hat{\beta}_{diff} = z_i(\beta - \hat{\beta}_{diff}) + f(x_i) + \varepsilon_i \cong f(x_i) + \varepsilon_i. \qquad (1.6.3)$$

If we apply conventional smoothing methods to these ordered pairs such as kernel estimation (see Section 3.2), then consistency, optimal rate of convergence results, and the construction of confidence intervals for f remain valid because $\hat{\beta}_{diff}$ converges sufficiently quickly to β that the approximation in the last part of (1.6.3) leaves asymptotic arguments unaffected. (This is indeed why we could apply the specification test after removing the *estimated* parametric effect.) Thus, in Figure 1.2 we have also plotted a nonparametric (kernel) estimate of f that can be compared with the quadratic estimate. In subsequent sections, we will elaborate this example further and provide additional ones.

1.7 Why Differencing?

An important advantage of differencing procedures is their simplicity. Consider once again the partial linear model $y = z\beta + f(x) + \varepsilon$. Conventional

Variable	Quadratic model		Partial linear model[a]	
	Coef	SE	Coef	SE
cust	−0.833	0.175	–	–
cust^2	0.040	0.009	–	–
wage	0.833	0.325	0.448	0.367
pcap	0.562	0.075	0.459	0.076
PUC	−0.071	0.039	−0.086	0.043
kwh	−0.017	0.089	−0.011	0.087
life	−0.603	0.119	−0.506	0.131
lf	1.244	0.434	1.252	0.457
kmwire	0.445	0.086	0.352	0.094
s_ε^2	.021		.018	
R^2	.618		.675	



[a] Test of quadratic versus nonparametric specification of scale effect: $V = n^{1/2}(s_{res}^2 - s_{diff}^2)/s_{diff}^2 = 81^{1/2}(.021 - .018)/.018 = 1.5$, where V is $N(0,1)$, Section 1.4.

Figure 1.2. Partial linear model – Log-linear cost function: Scale economies in electricity distribution.

4.6.2 Scale Economies in Electricity Distribution[7]

We now consider the example of Section 1.6 in considerably more detail. Suppose we have a slightly more general specification that is a semiparametric variant of the translog model (variable definitions may be found in Appendix E):

$$\begin{aligned} tc = {} & f(cust) + \beta_1 wage + \beta_2 pcap + \tfrac{1}{2}\beta_{11} wage^2 + \tfrac{1}{2}\beta_{22} pcap^2 \\ & + \beta_{12} wage \cdot pcap + \beta_{31} cust \cdot wage + \beta_{32} cust \cdot pcap + \beta_4 PUC \\ & + \beta_5 kwh + \beta_6 life + \beta_7 lf + \beta_8 kmwire + \varepsilon. \end{aligned} \tag{4.6.2}$$

Note that, in addition to appearing nonparametrically, the scale variable *cust* interacts parametrically with wages and the price of capital. One can readily verify that, if these interaction terms are zero (i.e., $\beta_{31} = \beta_{32} = 0$), then the cost function is homothetic. If in addition $\beta_{11} = \beta_{22} = \beta_{12} = 0$, then the model reduces to the log-linear specification of Section 1.6.

Differencing estimates of the parametric component of (4.6.2) are presented in Figure 4.5. (We use third-order optimal differencing coefficients, in which case $m = 3$.) Applying Proposition 4.5.2, we do not find significant statistical evidence against either the homothetic or the log-linear models. For example, the statistic testing the full version (4.6.2) against the log-linear specification, which sets five parameters to zero and is distributed χ^2_5 under the null, takes a value of 3.23. Estimates of nonprice covariate effects exhibit little variation as one moves from the full translog model to the homothetic and log-linear models.

The last column of Figure 4.5 contains HCSEs reported two ways: the first uses (4.5.9), which does not incorporate off-diagonal terms; the second uses (4.5.10), which does. We believe the latter to be more accurate here given the low order of differencing the small data set permits.

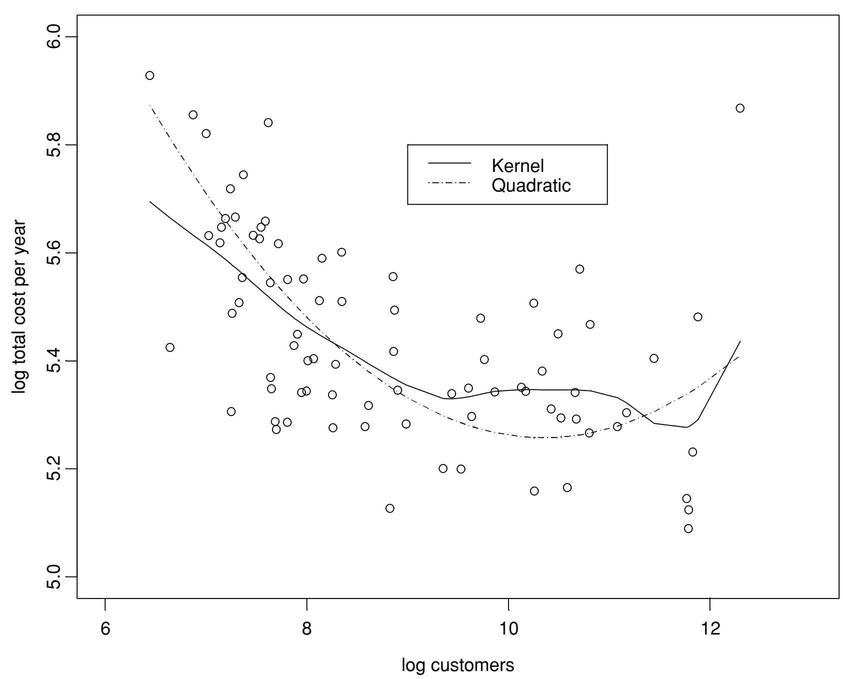
We may now remove the estimated parametric effect from the dependent variable and analyze the nonparametric effect. In particular, for purposes of the tests that follow, the approximation $y_i - z_i\hat{\beta} = z_i(\beta - \hat{\beta}) + f(x_i) + \varepsilon_i \cong f(x_i) + \varepsilon_i$ does not alter the large sample properties of the procedures. We use the estimates of the log-linear model to remove the parametric effect.

Figure 4.5 displays the ordered pairs $(y_i - z_i\hat{\beta}_{diff}, x_i)$ as well as a kernel estimate of *f*. Parametric null hypotheses may be tested against nonparametric alternatives using the specification test in Section 4.3. If we insert a constant function for f, then the procedure constitutes a test of significance of the scale variable x against a nonparametric alternative. The resulting statistic is 9.8, indicating a strong scale effect. Next we test a quadratic model for output. The resulting test statistic is 2.4, suggesting that the quadratic model may be inadequate.

[7] For a detailed treatment of these data, see Yatchew (2000).

Variable	Full model: semi-parametric translog		Homothetic model: semi-parametric homothetic		Log-linear model: semi-parametric Cobb–Douglas			
	Coef	SE	Coef	SE	Coef	SE	HCSE Eqn. 4.5.9	HCSE Eqn. 4.5.10
wage	−5.917	13.297	−6.298	12.453	0.623	0.320	0.343	0.361
pcap	−2.512	2.107	−1.393	1.600	0.545	0.068	0.078	0.112
$\frac{1}{2}\,wage^2$	0.311	2.342	0.720	2.130	–	–	–	–
$\frac{1}{2}\,pcap^2$	0.073	0.083	0.032	0.066	–	–	–	–
$wage \cdot pcap$	0.886	0.738	0.534	0.599	–	–	–	–
$cust \cdot wage$	0.054	0.086	–	–	–	–	–	–
$cust \cdot pcap$	−0.039	0.049	–	–	–	–	–	–
PUC	−0.083	0.039	−0.086	0.039	−0.075	0.038	0.034	0.033
kwh	0.031	0.086	0.033	0.086	0.008	0.086	0.074	0.089
life	−0.630	0.117	−0.634	0.115	−0.628	0.113	0.095	0.097
lf	1.200	0.450	1.249	0.436	1.327	0.434	0.326	0.304
kmwire	0.396	0.087	0.399	0.087	0.413	0.084	0.090	0.115
s_ε^2	.01830		.0185		.01915			
R^2	.668		.665		.653			

Estimated scale effect



Test of full (translog) model versus log-linear (Cobb–Douglas) model: χ_5^2 under H_o : 3.23. Test of quadratic versus nonparametric specification of scale effect: $V = (mn)^{1/2}(s_{res}^2 - s_{diff}^2)/s_{diff}^2 = (3*81)^{1/2}(.0211 - .0183)/.0183 = 2.4$ where V is $N(0,1)$. Kernel estimate produced using *ksmooth* function in *S-Plus*. The last two columns of the table contain heteroskedasticity-consistent standard errors (HCSEs).

Figure 4.5. Scale economies in electricity distribution.

P1: FZZ/FFX P2: FZZ/FFX QC: FZZ
CB541-Yatchew CB541-04 March 3, 2003 18:45

To provide further illustrations of differencing procedures we divide our data into two subpopulations: those that deliver additional services besides electricity, that is, public utility commissions (PUC), and those that are pure electricity distribution utilities (non-PUC). The numbers of observations in the two subpopulations are $n_{PUC} = 37$ and $n_{nonPUC} = 44$. We denote differencing estimates of parametric effects and of residual variances as $\hat{\beta}_{PUC}$, $\hat{\beta}_{nonPUC}$, s^2_{PUC}, and s^2_{nonPUC}. For each subpopulation, we estimate the log-linear model using the differencing estimator and report the results in Figure 4.6.

To test whether PUC and non-PUC entities experience the same parametric effects, we use

$$(\hat{\beta}_{PUC} - \hat{\beta}_{nonPUC})'\left(\hat{\Sigma}_{\hat{\beta}_{PUC}} + \hat{\Sigma}_{\hat{\beta}_{nonPUC}}\right)^{-1} \times (\hat{\beta}_{PUC} - \hat{\beta}_{nonPUC}) \xrightarrow{D} \chi^2_{dim(\beta)}. \tag{4.6.3}$$

The computed value of the χ^2_6 test statistic is 6.4, and thus the null is not rejected. Next, we constrain the parametric effects to be equal across the two types of utilities while permitting distinct nonparametric effects. This is accomplished by taking a weighted combination of the two estimates

$$\hat{\beta}_{weighted} = \left[\hat{\Sigma}^{-1}_{\hat{\beta}_{PUC}} + \hat{\Sigma}^{-1}_{\hat{\beta}_{nonPUC}}\right]^{-1}\left[\hat{\Sigma}^{-1}_{\hat{\beta}_{PUC}} \cdot \hat{\beta}_{PUC} + \hat{\Sigma}^{-1}_{\hat{\beta}_{nonPUC}} \cdot \hat{\beta}_{nonPUC}\right] \tag{4.6.4}$$

with estimated covariance matrix

$$\hat{\Sigma}_{\hat{\beta}_{weighted}} = \left[\hat{\Sigma}^{-1}_{\hat{\beta}_{PUC}} + \hat{\Sigma}^{-1}_{\hat{\beta}_{nonPUC}}\right]^{-1}. \tag{4.6.5}$$

The results are reported in Table 4.3.[8] The data can be purged of the estimated parametric effects, and separate nonparametric curves can be fitted to each

[8] A numerically similar estimator with the same large sample properties may be constructed by differencing the data within each subpopulation and then stacking as follows

$$\begin{bmatrix} Dy_{PUC} \\ Dy_{nonPUC} \end{bmatrix} = \begin{bmatrix} DZ_{PUC} \\ DZ_{nonPUC} \end{bmatrix}\beta + \begin{bmatrix} Df_{PUC}(x_{PUC}) \\ Df_{nonPUC}(x_{nonPUC}) \end{bmatrix} + \begin{bmatrix} D\varepsilon_{PUC} \\ D\varepsilon_{nonPUC} \end{bmatrix}.$$

Let $\hat{\beta}$ be the OLS estimator applied to the preceding equation. Then, the common residual variance may be estimated using

$$s^2 = \frac{1}{n}\left(\begin{bmatrix} Dy_{PUC} \\ Dy_{nonPUC} \end{bmatrix} - \begin{bmatrix} DZ_{PUC} \\ DZ_{nonPUC} \end{bmatrix}\hat{\beta}\right)'\left(\begin{bmatrix} Dy_{PUC} \\ Dy_{nonPUC)} \end{bmatrix} - \begin{bmatrix} DZ_{PUC} \\ DZ_{nonPUC} \end{bmatrix}\hat{\beta}\right),$$

and the covariance matrix of $\hat{\beta}$ may be estimated using

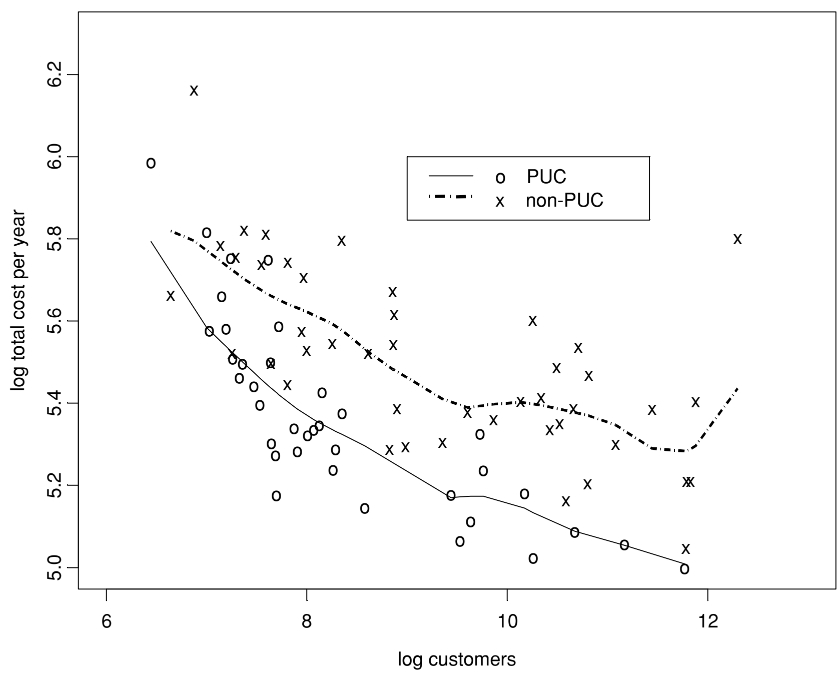
$$\hat{\sum}_{\hat{\beta}} = \left(1 + \frac{1}{2m}\right)\frac{s^2}{n}\left[(DZ_{PUC})'(DZ_{PUC}) + (DZ_{nonPUC})'(DZ_{nonPUC})\right]^{-1},$$

where m is the order of (optimal) differencing.

P1: FZZ/FFX P2: FZZ/FFX QC: FZZ
CB541-Yatchew CB541-04 March 3, 2003 18:45

Variable	Partial linear model[a]			
	PUC		non-PUC	
	Coef	SE	Coef	SE
wage	0.65	0.348	1.514	0.684
pcap	0.424	0.090	0.632	0.113
kwh	0.108	0.121	0.079	0.123
life	−0.495	0.131	−0.650	0.199
lf	1.944	0.546	0.453	0.702
kmwire	0.297	0.109	0.464	0.123
s_ε^2	0.013		0.023	

Estimated scale effect



[a] Order of differencing $m = 3$.

Figure 4.6. Scale economies in electricity distribution: PUC and non-PUC analysis.

Table 4.3. *Mixed estimation of PUC/non-PUC effects: Scale economies in electricity distribution.*[a]

Variable	Coef	SE
wage	0.875	0.304
pcap	0.526	0.067
kwh	0.066	0.086
life	−0.547	0.107
lf	1.328	0.422
kmwire	0.398	0.078

[a]Estimates of parametric effects are obtained separately for PUC and non-PUC subpopulations. Hence, no PUC effect is estimated. The estimates above are obtained using (4.6.4) and (4.6.5).

subset of the data, as in the bottom panel of Figure 4.6. The PUC curve lies below the non-PUC curve consistent with our earlier finding that PUC entities have lower costs (see PUC coefficients in Figure 4.5).

We may now adapt our test of equality of regression functions in Section 4.4.2 to test whether the curves in Figure 4.6 are parallel, that is, whether one can be superimposed on the other by a vertical translation. This may be accomplished simply by removing the mean of the purged dependent variable from each of the two subpopulations.

Define the within estimate to be the weighted average of the subpopulation variance estimates, keeping in mind that the estimated parametric effect has been removed using, say, $\hat{\beta}_{weighted}$:

$$s_w^2 = \frac{n_{PUC}}{n} s_{PUC}^2 + \frac{n_{nonPUC}}{n} s_{nonPUC}^2. \tag{4.6.6}$$

Let y_{PUC}^{purge} be the vector of data on the dependent variable for PUCs with the estimated parametric effect removed and then centered around 0 and define y_{nonPUC}^{purge} similarly.[9] Now stack these two vectors and the corresponding data on the nonparametric variable x to obtain the ordered pairs (y_i^{purge}, x_i) $i = 1, \ldots, n$. Let P_p be the permutation matrix that reorders these data so that the nonparametric variable x is in increasing order. Note that, because separate equations

[9] Because the hypothesis that the parametric effects are the same across the two populations has not been rejected, one may use subpopulation estimates β_{PUC}^2 and β_{nonPUC}^2 or the weighted estimate $\hat{\beta}_{weighted}$ when computing s_{PUC}^2, s_{nonPUC}^2, and s_p^2.

were estimated for the two subpopulations, z does not contain the PUC dummy. Define

$$s_p^2 = \frac{1}{n} y^{purge'} P_p' D' D P_p y^{purge}. \tag{4.6.7}$$

If the null hypothesis is true, then differencing will still remove the nonparametric effect in the pooled data and s_p^2 will converge to σ_ε^2. Otherwise, it will generally converge to some larger value. Applying Proposition 4.4.2 with $m = 1$, we obtain a value of 1.77 for $\Upsilon / s_w^2 (2\hat{\pi}_\Upsilon)^{1/2}$, which, noting that this is a one-sided test, suggests that there is some evidence against the hypothesis that the scale effects are parallel. Finally, we note that, given the size of the two subsamples, one must view the asymptotic inferences with some caution. An alternative approach that generally provides better inference in moderately sized samples would be based on the bootstrap, which is discussed in Chapter 8.

4.6.3 Weather and Electricity Demand

In a classic paper, Engle et al. (1986) used the partial linear model to study the impact of weather and other variables on electricity demand. We estimate a similar model in which weather enters nonparametrically and other variables enter parametrically. Our data consist of 288 quarterly observations in Ontario for the period 1971 to 1994. The specification is

$$elec_t = f(temp_t) + \beta_1 relprice_t + \beta_2 gdp_t + \varepsilon, \tag{4.6.8}$$

where *elec* is the log of electricity sales, *temp* is heating and cooling degree days measured relative to 68 °F, *relprice* is the log of the ratio of the price of electricity to the price of natural gas, and *gdp* is the log of gross provincial product. We begin by testing whether electricity sales and *gdp* are cointegrated under the assumption that *relprice* and *temp* are stationary (setting aside issues of global warming). The Johansen test indicates a strong cointegrating relationship. We therefore reestimate the model in the form

$$elec_t - gdp_t = f(temp_t) + \beta_1 relprice_t + \varepsilon. \tag{4.6.9}$$

Figure 4.7 contains estimates of a pure parametric specification for which the temperature effect is modeled using a quadratic as well as estimates of the partial linear model (4.6.9). The price of electricity relative to natural gas is negative and quite strongly significant. In the partial linear model, the