Queen's Economics Department Working Paper No. 257

Seasonality in Regression: An Application of Smoothness Priors

Mark Gersovitz
Princeton University

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-1977

Seasonality in Regression: An Application of Smoothness Priors

Mark Gersovitz

Department of Economics
Princeton University
Princeton, NJ 08540
U.S.A.

and

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

Abstract

This article argues that conventional approaches to the treatment of seasonality in econometric investigation are often inappropriate. A more appropriate technique is to allow all regression coefficients to vary with the season, but to constrain them to do so in a smooth fashion. A Bayesian method of estimating smoothly varying seasonal coefficients is developed, based on Shiller's (1973) approach to estimating distributed lags. In a sampling experiment, this technique outperforms ordinary least squares by a substantial margin. An application of this technique to the estimation of the demand for soft drinks is also presented.

The authors would like to thank R. J. Arnott, C. M. Beach, P. T. Chinloy, J. W. Eaton, M. C. Lovell, three referees, and R. E. Quandt for helpful comments on earlier drafts. This paper was published in the *Journal of the American Statistical Association*, 73, 1978, 264–273.

January, 1977

1. Introduction

Many economic time series display regular seasonal fluctuations. If the seasonal fluctuations in the independent variables fully accounted for the seasonal fluctuations in the dependent variables, no problem would exist for econometricians. Indeed, by imparting additional variation to the independent variables, seasonal fluctuations would increase the precision of coefficient estimates. It is often the case, however, that when seasonally varying dependent variables are regressed on seasonally varying independent variables, the resulting residuals have a seasonal pattern. Faced with this situation, the practicing econometrician typically does one of two things: Either he inserts dummy variables to capture the effects of seasonality, or he employs seasonally adjusted data (usually provided by official sources). In this article, a third approach is advocated. We propose to allow all of the parameters of the model to vary from season to season, but to constrain them to do so in a smooth fashion.

The rationale for this approach is developed in this section. An estimation procedure which implements the approach is presented in Section 2. The results of some sampling experiments which investigate the performance of this procedure are given in Section 3. Finally, in Section 4, the technique is applied to the estimation of the demand for soft drinks in Canada.

The economic model which explains an economic time series $x(t)$ may be written as

$$x(t) = f\big(\boldsymbol{z}(t), \boldsymbol{u}(t), \boldsymbol{\theta}\big), \tag{1.1}$$

where $\boldsymbol{z}(t)$ is a vector of other economic time series, which may include lagged values, including lagged values of $x(t)$ itself, $\boldsymbol{u}(t)$ is a vector of random errors, and $\boldsymbol{\theta}$ is a vector of parameters. In some cases, it may be possible to rewrite (1.1) as

$$x(t) = g\big(\boldsymbol{z}(t), \boldsymbol{u}(t), \boldsymbol{\theta}_1\big) + S(t, \boldsymbol{\theta}_2). \tag{1.2}$$

That is, it may be possible to break $x(t)$ into two parts, one of which (S) varies systematically with the season but does not depend on $\boldsymbol{z}(t)$, and the other of which (g) does not show any systematic seasonal variation. Of course, $x(t)$ may be the logarithm of a series, so that (1.2) would actually be a multiplicative relationship. If the model can be written in this way, we say that it displays separable seasonality; if not, we say that it displays inseparable seasonality. If deseasonalization of the time series $x(t)$ is to be practical, the model underlying x must display separable seasonality. Otherwise, one could not hope to estimate $\boldsymbol{\theta}_2$ independently of $\boldsymbol{\theta}_1$, and hence a deseasonalized series could not be arrived at without first estimating (1.1).

Many techniques for deseasonalization have been proposed; see, among others, Lovell (1963), Jorgenson (1964), Shiskin, Young, and Musgrave (1967), and Sims (1974). It should be noted that inserting seasonal dummies into a regression on raw data is equivalent to using one of these standard deseasonalization techniques; see Lovell (1963). All such techniques attempt to estimate the seasonal component without estimating, or even knowing the form of, equation (1.1). Thus they all assume that seasonality is separable.

Economic theory, however, provides no support for this assumption. The one unifying principle of neoclassical theory is that market behavior should be derived from underlying tastes and technology by assuming that agents maximize some quantity such as utility or profits. Thus seasonal variation in market variables should be derived from seasonal influences on tastes and technology. It is easy to construct simple models of this type, and we have investigated a number of them. In every case where the utility, cost, or production functions were realistic, the demand or supply functions derived from them displayed inseparable seasonality. Thus we believe that inseparable seasonality is the rule rather than the exception for economic time series.

In principle, the econometrician should use economic theory to specify the functional forms of the relationships he or she estimates, including the ways in which seasonal factors enter into them. In practice, however, partly because most theory applies to individual agents while most time series data apply to broad aggregates, microeconomic theory provides little guidance about functional forms. The econometrician typically falls back on some variant of the general linear model to serve as an approximation of the true functional form which is complicated and unknown. If seasonality is not separable, this approximation should be different in every season, so that every parameter is allowed to vary from season to season. Thus an appropriate model to estimate is

$$y_t = \sum_{i=1}^{\lambda} D_{it} \boldsymbol{X}_t \boldsymbol{\beta}_i + \varepsilon_t, \tag{1.3}$$

where y_t is a dependent variable and $\boldsymbol{X}_t$ is a vector of independent variables at time t, D_{it} is a scalar equal to unity if t is in the i^{th} season and equal to zero otherwise, $\boldsymbol{\beta}_i$ is a vector of coefficients for season i, and the number of seasons is λ.

The model (1.3) is simply a generalization of the traditional dummy variable procedure in which all parameters, not merely the intercept, are allowed to vary seasonally. Hence it can approximate a true functional form in which seasonality is inseparable. Ordinary least squares estimation of (1.3) is equivalent to simply fitting a separate relationship for every season, and thus it is likely to lead to rather imprecise estimates unless the sample size is large. One would, therefore, like to impose some additional restrictions on the model. In many situations, it may be reasonable to impose the restriction that the values of each parameter in adjacent seasons will not differ greatly. We refer to this as the assumption of smooth seasonality. It is often a useful assumption, and in the next section we present a technique for its econometric implementation. First, however, we examine the smooth seasonality hypothesis more closely.

Three objections may be raised to the assumption that parameters vary smoothly from season to season. First, such an assumption is simply inappropriate in some situations. For example, the demand for Easter eggs varies seasonally, but the variation is probably not very smooth, and, moreover, it will be different in different years. In such situations, the smooth seasonality assumption should not be employed.

Second, the smooth seasonality hypothesis puts no restrictions on relationships among the coefficients on different variables. Unfortunately, such restrictions are not, in general, justified. Consider the utility function

$$U(x_1, sx_2), \tag{1.4}$$

where x_1 and x_2 are the quantities of goods 1 and 2, and s is a parameter which varies smoothly with the season. If the functional form of U were unknown, one would often estimate log-linear relationships between quantities demanded, prices, and income, in which the parameters of interest are price and income elasticities. It can straightforwardly be shown that when demand functions are derived from (1.4), the effect of a change in s on the income elasticity of the demand for each good is quite different from the effect of such a change on the price elasticities. Thus the relationships among coefficients on different variables can be expected to vary with the season.

A final objection is that phenomena which are usually called seasonal are often directly linked to one or more dimensions of the weather, such as temperature or rainfall. Thus one should incorporate weather directly into the model to explain seasonality. Such a proposal is attractive in principle, but it may be very hard to implement. Weather has many dimensions and often varies enormously across regions for which statistical data are available. Thus it is doubtful that seasonality could be adequately explained by only a few weather variables. Even if it could be, it seems likely that the parameters of the model would depend on the weather variable(s) in a complicated and nonlinear way, so that specifying the model might be very difficult. Moreover, some seasonal phenomena may depend not so much on actual weather as on the weather that usually occurs in those seasons, so that use of weather data may introduce an errors-in-variables problem. The hypothesis of smoothly varying seasonal parameters may be regarded as a way of approximating the effects of average weather on the model's parameters. In Section 4, where we present an example of our smoothness technique, we also examine methods for the explicit use of weather data.

2. Seasonally Varying Parameters and Smoothness Priors

The problem of estimating parameters which vary with the season is not unlike the problem of estimating a distributed lag model. In both cases, the main difficulty is the large number of parameters that could potentially be estimated, and the obvious solution is to constrain the estimates in some way, The most elegant and flexible technique for constraining the coefficients of a distributed lag is the Bayesian technique recently proposed by Shiller (1973). In this section, his technique is adapted to the case of seasonally varying parameters; where possible, Shiller's notation is used.

A natural constraint to impose is that the coefficients of a given variable should vary smoothly across the seasons. Depending on what is meant by "smoothly," this requirement could imply that the first, second, or higher differences should be small. The technique proposed here could be used to impose any of these constraints. To conserve space, only first degree smoothness priors (which impose constraints on the second differences) are considered here.

For simplicity, consider the case with only one independent variable. Suppose that

$$y_t = \sum_{i=1}^{\lambda} \beta_i D_{it} x_t + \varepsilon_t, \tag{2.1}$$

where x_t and y_t are scalar time series at time t, $D_{it} = 1$ if t is in the i^{th} season, and $D_{it} = 0$ otherwise. For a first degree smoothness prior, second differences between the seasonally varying coefficients are assumed to be small. That is

$$(\beta_i - \beta_{i-1}) - (\beta_{i-1} - \beta_{i-2}) = \beta_i - 2\beta_{i-1} + \beta_{i-2} \tag{2.2}$$

is assumed to be small for all periods i. Note that the periodicity of the seasons implies that the first coefficient is linked to the last as well as to the second; thus $\beta_{i-1} = \beta_\lambda$ if $i = 1$, and so on.

The second differences of the vector $\boldsymbol{\beta}$ may be written as

$$\boldsymbol{u} = \boldsymbol{R}_1 \boldsymbol{\beta}, \tag{2.3}$$

where $\boldsymbol{R}_1$ is a $\lambda \times \lambda$ matrix which generates (2.2). For example, when $\lambda = 4$ (i.e., when the data are quarterly),

$$\boldsymbol{R}_1 = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -2 \\ -2 & 1 & 0 & 1 \end{bmatrix}. \tag{2.4}$$

In Shiller's development, the prior information that the second differences are small is represented by assuming that the second differences are normally and independently distributed with mean zero and variance ζ^2. Hence $\boldsymbol{u}$ comes from a spherical normal distribution with covariance matrix $\zeta^2 \mathbf{I}$. In the case of seasonally varying coefficients, however, this simple formulation is inadmissible, because the second differences are linearly dependent. This fact is easily verified by noting that the first $\lambda - 1$ rows of $\boldsymbol{R}_1$ sum to minus the λ^{th} row.

Since the second differences of $\boldsymbol{\beta}$ are linearly dependent, the prior information that they are small presumably cannot be represented by the assumption that they are normally and independently distributed (but see the following discussion). Instead, we specify that they are normally distributed with a variance-covariance matrix which is consistent with their being linearly dependent. The simplest formulation for the variance-covariance matrix of $\boldsymbol{u}$ which takes into account the particular form of this linear dependence is

$$\zeta^2 \boldsymbol{\Omega} = \zeta^2 \begin{bmatrix} 1 & \omega & \omega & \cdots & \omega & \omega \\ \omega & 1 & \omega & \cdots & \omega & \omega \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \omega & \omega & \omega & \cdots & \omega & \omega \\ \omega & \omega & \omega & \cdots & \omega & 1 \end{bmatrix}, \tag{2.5}$$

where $\omega = -1/(\lambda - 1)$, and $\boldsymbol{\Omega}$ is $\lambda \times \lambda$. Since $|\boldsymbol{\Omega}| = 0$, $\boldsymbol{u}$ has a degenerate distribution. Therefore, from now on, we deal with the distribution of $\boldsymbol{u}^*$, which is equal to $\boldsymbol{u}$ with the last element deleted. Define $\boldsymbol{R}_1^*$ as a $(\lambda - 1) \times \lambda$ matrix consisting of $\boldsymbol{R}_1$ with the last row deleted. Then

$$\boldsymbol{u}^* = \boldsymbol{R}_1^* \boldsymbol{\beta}, \tag{2.6}$$

and $\boldsymbol{u}^*$ is normally distributed with covariance matrix $\zeta^2 \boldsymbol{\Omega}^*$, where $\boldsymbol{\Omega}^*$ is derived by deleting the last row and last column of $\boldsymbol{\Omega}$. Finally, define

$$\tilde{\boldsymbol{R}}_1 = \boldsymbol{\Omega}^{*\,-1/2} \boldsymbol{R}_1^* \tag{2.7}$$

and

$$\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{R}}_1 \boldsymbol{\beta}. \tag{2.8}$$

The vector $\tilde{\boldsymbol{u}}$ has a spherical normal distribution with covariance matrix $\zeta^2 \mathbf{I}_{\lambda - 1}$.

In order to impose a prior on $\tilde{\boldsymbol{u}}$, it is necessary to know $\boldsymbol{\Omega}^{*\,-1/2}$ From the form of $\boldsymbol{\Omega}$, it is obvious that this must be a matrix with α_1 on the principal diagonal and α_2 everywhere else. Solving for α_1 and α_2 is straightforward. There are two solutions, one of which is

$$\alpha_1 = \frac{\lambda + \lambda^{1/2} - 2}{(\lambda - 1)^{1/2} \lambda^{1/2}}, \quad \alpha_2 = \frac{\lambda^{1/2} - 1}{(\lambda - 1)^{1/2} \lambda^{1/2}}. \tag{2.9}$$

This solution may be derived from a formula provided by Nerlove (1971).

The matrix $\tilde{\boldsymbol{R}}_1$ plays exactly the same role as Shiller's $\boldsymbol{R}_1$. Thus combining the prior on $\boldsymbol{u}$ with a normal likelihood function for the ε_t, which are assumed to be n.i.d. with variance σ^2, yields a posterior distribution which is normal. The smoothness estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}, \tag{2.10}$$

where

$$\tilde{\boldsymbol{X}} \equiv \begin{bmatrix} \boldsymbol{D}_1 \boldsymbol{x} \;\; \boldsymbol{D}_2 \boldsymbol{x} \cdots \boldsymbol{D}_\lambda \boldsymbol{x} \\ k \tilde{\boldsymbol{R}}_1 \end{bmatrix}, \tag{2.11}$$

$$\tilde{\boldsymbol{y}} \equiv \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}, \tag{2.12}$$

and

$$k \equiv \sigma / \zeta. \tag{2.13}$$

Obtaining $\hat{\boldsymbol{\beta}}$ is considerably easier if one replaces $k\tilde{\boldsymbol{R}}_1$ in (2.11) by

$$\left(\frac{\lambda - 1}{\lambda}\right)^{1/2} k \boldsymbol{R}_1 \tag{2.14}$$

and lengthens the vector of zeros in $\tilde{\boldsymbol{y}}$ accordingly. It is proved in Appendix 1 that $\hat{\boldsymbol{\beta}}$ is unaffected by this substitution.[1] It is easy to obtain $\hat{\boldsymbol{\beta}}$ using an ordinary least-squares regression package; it is merely necessary to add λ zeros to the vector $\boldsymbol{y}$, and the λ rows of $\boldsymbol{R}_1$, multiplied by

$$\big((\lambda-1)/\lambda\big)^{1/2}k,$$

to the $\boldsymbol{X}$ matrix. This procedure can thus be interpreted in the context of mixed estimation, in a manner analogous to Taylor's (1974) interpretation of Shiller's procedure. If k is large, the dummy observations will carry a lot of weight, and as a result the estimated second differences will be small. The choice of k will be discussed in Section 3.

The procedure outlined above can be extended directly to the case where there is more than one independent variable. A different prior must be specified for each set of seasonally varying coefficients, and hence a different k must be used for each. If there are T observations and n variables, all with seasonally varying coefficients, then $\tilde{\boldsymbol{X}}$ will have $n\lambda$ columns and $T+n\lambda$ rows, with λ rows of dummy observations for each of the n variables.

3. Sampling Experiments

Several sampling experiments were performed to investigate the performance of the smooth seasonality technique. The main objectives of these experiments were to see how the smoothness technique compares with ordinary least squares, and to find out how the choice of k affects the estimates.

The model examined is

$$y_t = \sum_{i=1}^{12} a_i D_{it} + \sum_{i=1}^{12} b_i D_{it} x_t + u_t, \tag{3.1}$$

where D_{it} is a dummy variable that equals one when t equals i plus an integer multiple of 12, and that equals zero elsewhere. The error term u_t is normally distributed with mean zero and variance σ^2. The independent variable x_t is generated by

$$x_t = \sum_{i=1}^{12} c_i D_{it}(Ar^t) + e_t, \tag{3.2}$$

where e_t has mean zero and variance σ_e^2. Thus x_t trends upward but also varies seasonally (due to the c_i) and randomly.

[1] We are grateful to an anonymous referee for pointing this out to us. It may also be noted that our procedure is equivalent to using an Aitken estimator on the complete stacked regression, with the Moore-Penrose generalized inverse of the variance-covariance matrix $\boldsymbol{\Omega}$.

In the sampling experiments reported here, the following parameter values were used:

$$
\begin{aligned}
&c_1 = c_2 = c_3 = 0.9,\\
&c_4 = c_5 = c_6 = c_{10} = c_{11} = c_{12} = 1.0,\\
&c_7 = c_8 = c_9 = 1.1,\\
&A = 10,\, r = 1.005,\, \sigma_e = 1,\\
&a_1 = 8,\, a_2 = 8,\, a_3 = 9,\, a_4 = 10a_5 = 10,\, a_6 = 11,\\
&a_7 = 12,\, a_8 = 12,\, a_9 = 11,\, a_{10} = 10,\, a_{11} = 10,\, a_{l2} = 9,\\
&b_1 = 1.173205,\, b_2 = 1.1,\, b3 = 1.0,\, b_4 = 0.9,\, b_5 = 0.826795,\, b_6 = 0.8,\\
&b_7 = 0.826795,\, b_8 = 0.9,\, b_9 = 1.0,\, b_{10} = 1.1,\, b_{11} = 1.173205,\, b_{12} = 1.2
\end{aligned}
\tag{3.3}
$$

Note that the b_i follow a sinusoidal pattern. They were in fact generated by the equation

$$b_i = 1.0 + 0.2\cos(\pi i/6).$$

The pattern of the a_i, though regular, is somewhat less smooth.

The correct values of k_a and k_b are σ/ζ_a and σ/ζ_b, respectively. But although σ is known, ζ_a and ζ_b, the standard deviations of the priors on the second differences of the a_i and the b_i, are not. Since the assumed a_i and b_i were not actually generated as realizations of multivariate normal distributions on their second differences, their standard deviations are not good estimators of ζ_a and ζ_b. Shiller suggests that, for the distributed lag case, ζ should be derived by assuming that the lag has a vee shape and the sum of coefficients expected by the investigator. An analogous procedure for the seasonal case is to assume that the seasonal coefficients follow a vee wave, with the true (or expected) amplitude, rising linearly for half the year and falling linearly for the other half. One could also assume that the coefficients lie on a square wave with given amplitude, so that they are equal to their largest value for half the year and to their smallest for the other half. Given some such set of assumed β_i, one may compute the value of ζ using

$$\zeta^2 = \frac{\boldsymbol{\beta}^\top \boldsymbol{R}_1^{*\top} \boldsymbol{\Omega}^{*-1} \boldsymbol{R}_1^* \boldsymbol{\beta}}{\lambda - 1}, \tag{3.4}$$

or, equivalently from Appendix 1,

$$\zeta^2 = \boldsymbol{\beta}^\top \boldsymbol{R}_1{}^\top \boldsymbol{R}_1 \boldsymbol{\beta}/\lambda. \tag{3.5}$$

Expression (3.4) has the form of a maximum likelihood estimate of ζ^2, and (3.5) is simply the mean of the squared second differences of the elements of $\boldsymbol{\beta}$.

If the a_i follow a vee wave with their actual amplitude, the resulting value of ζ_a is 0.544331, and if they follow a square wave, the resulting value of ζ_a is 2.309401; similar assumptions on the b_i, which have one-tenth the amplitude, yield values of ζ_b one-tenth as large. These values of ζ_a and ζ_b were used in the sampling experiments.

The square wave assumption yields relatively small values of k, and the vee wave assumption yields relatively large values of k, so the two estimators will be referred to as the small-k and large-k estimators, respectively.

The results from two sampling experiments are presented in Tables 1 and 2. In both cases, the model was given by (3.1), (3.2), and (3.3), the number of observations was 120 (corresponding to ten years of monthly data), and the number of replications was 100. These simulations were written in FORTRAN on a Burroughs B6700, using 48-bit floating point arithmetic. The error terms were generated by a routine which approximates a normally distributed random variate by the sum of twelve pseudo-random uniform variates. In the two experiments reported on here, the standard deviation of the error term in (3.1), σ, is 0.5 and 1.0, respectively. Different values of σ were investigated because, as σ increases, the weight accorded the prior information also increases.

Root mean square errors (RMSE) are presented in columns one to three of Tables 1 and 2. Note that $\bar{a}$ refers to the average of the a_i, so that the entries in columns 1 to 3 beside $\bar{a}$ refer to the RMSE of the average, while the entries beside "avg." refer to the average of the RMSEs over the 12 periods. It is evident from Table 1 that the smoothness estimators have substantially lower RMSEs than ordinary least-squares (OLS) estimators; the only coefficients for which the difference is not substantial are $\bar{a}$ and $\bar{b}$. The large-k RMSEs are somewhat smaller than the small-k RMSEs; however, the former are much more variable than the latter. The results in Table 2 confirm those in Table 1, the main difference being that all RMSEs are substantially larger, and the relative performance of OLS is worse.

Mean biases are presented in columns four through six of Tables 1 and 2. An asterisk indicates that bias was significant at the 5 percent level, according to a simple nonparametric test on the number of replications for which the estimate exceeded the true value; critical points were 39 and 61 replications, using the normal approximation to the binomial. The small-k estimates exhibit some bias, but it is not large and mainly affects a few coefficients. The large-k estimates, on the other hand, exhibit very substantial bias, especially for $\sigma = 1.0$. The estimates apparently tend to reduce the amplitude of the true seasonal coefficients; large coefficients are underestimated, and small ones are overestimated.

Experiments for larger values of σ, not reported here, confirm the results suggested by comparing Tables 1 and 2: As the standard error of the regression increases, the RMSEs of the smoothness estimates increase. When σ is very large, the large-k estimates exhibit almost no seasonal variation at all, and the small-k estimates are substantially biased.

It should be remembered that, in all these experiments, k was set equal to $\sigma/zeta$, so that the weights on the dummy observations increased linearly with σ. It would appear to be the case that, in order to avoid excessive bias, k should vary less than proportionately with σ.

4. An Application

To illustrate how the smooth seasonality technique performs in practice, it was applied to the estimation of the demand for soft drinks in Canada, using monthly data. The model estimated is

$$C_t = \sum_{i=1}^{12} D_{it}(a_{1i}S_t + a_{2i}SDP_t + a_{3i}FP_t + a_{4i}EX_t + a_{5i}Y_t) + u_t, \tag{4.1}$$

where C_t is the log of per capita soft drink consumption, $S_t = 1$, SDP_t is the log of the price of soft drinks relative to a nonfood price index, FP_t is the log of the price of food relative to the nonfood price index, EX_t is the log of per capita real expenditure, and Y_t is a variable representing the relative proportion of young people in the population. The index i is one in January. Monthly data for 1959:1 to 1974:6 are employed, so that there are 186 observations. Since all variables are measured in natural logarithms, the coefficients are elasticities. Precise definitions of all variables are given in Appendix 2. This model is presented here as an illustration of technique, not as a definitive analysis of soft drink demand.

The values of k_1 through k_5 for this example were chosen as follows. First, OLS was applied to equation (4.1). The ζ_j were then computed from (3.5) on the assumption that the j^{th} coefficients had mean equal to the mean of the OLS coefficients and followed a square wave which varied from +100 percent to −100 percent of that mean. The quantity k_j was then computed as $\hat{\sigma}/\zeta_j$, where $\hat{\sigma}$ is the estimated standard error from the OLS regression. This procedure yielded the following values for the k_j:

$$k_1 = 0.0111, \;\; k_2 = 0.0822, \;\; k_3 = 0.2466, \;\; k_4 = 0.0123, \;\; k_5 = 0.1206.$$

For purposes of comparison, two more restrictive models were also estimated. One, which will be referred to as "Lovell", because it uses the dummy-intercept treatment of seasonality dealt with by Lovell (1963), constrains a_{ji} to equal a_j for all variables except the intercept. The second, which will be referred to as "Unified", constrains a_{ji} to equal a_j for all variables, and thus takes no account of seasonality at all.

The results of smoothness estimation of (4.1) with the k_j given above, of OLS estimation of (4.1), and of the Lovell and Unified models, are presented in Table 3. Numbers in parentheses are t statistics. The Unified model is clearly inappropriate. Its estimates are wildly different from those of the other three models, its R^2 is very low, and an F test of Unified against OLS rejects the former at all normal significance levels. The Lovell model is not as clearly inappropriate; OLS fits significantly better than it does at the ten percent level, but not at the five percent level. This does suggest, however, that it would be dangerous to accept the Lovell model as a complete treatment of seasonality in soft drink demand. There is no evidence of twelfth-order autocorrelation for either the Lovell or Smoothness models, according to a regression of the residuals on those lagged twelve months. In contrast, Unified displays evidence of severe twelfth-order autocorrelation (significant at more than the .01 level), suggesting that this test on the residuals is a useful diagnostic.

In terms of the estimates of average coefficients, there is little to choose between the Smoothness, OLS, and Lovell approaches. However, the Lovell approach necessarily gives no information about the seasonal pattern of the coefficients, and the OLS estimates jump around so much that they also provide no useful information. The Smoothness estimates, on the other hand, generally vary from month to month in a simple, regular fashion. Thus, if the investigator has any interest in the seasonal pattern of the coefficients, either because it is interesting in itself or because it may provide evidence of misspecification (if, for example, the peaks are in the wrong season), the smoothness technique would appear to be well worth employing.

One way to test the appropriateness of the smoothness restrictions is to use Theil's test for the compatibility of prior and sample information (Theil 1971, pp. 350–351). Under the null hypothesis, the test statistic is distributed as a chi-squared random variable with 55 degrees of freedom. The value of the statistic for the prior we used is 197.9, which is more than twice the 0.005 tail value. Thus the sample information appears to be inconsistent with the smoothness prior. It should be noted, however, that if the k_j are reduced by a factor of two, the resulting priors, which seemed to us too weak, do pass the Theil test at the .05 level.

As discussed in Section 2, an alternative to the seasonally varying parameters model is the explicit introduction of weather into the regression. The dimension of weather most relevant to soft drink demand is temperature. Accordingly, we introduced a temperature variable, W_t, which is a weighted average of daily maximum temperatures in Canada's three largest metropolitan areas, in or near which most Canadians live; see Appendix 2.

Initially, we simply added W_t to the specification, either additively or multiplicatively. The results were disappointing. Neither W_t nor W_t times each of the five other variables added significantly to the explanatory power of OLS applied to (4.1). Although they did add significantly to the power of the Unified model, the resulting equations were not very impressive.

We then experimented with a spline approach (Poirier, 1975) to allow for nonlinear effects. We felt that 35° F might represent a temperature below which changes in temperature would have no effect on the demand for soft drinks, while 70° F might represent a temperature beyond which changes in temperature would greatly affect demand. Accordingly, we defined

$$W_1 = \begin{cases} W - 35 & \text{if } W \geq 35 \\ 0 & \text{otherwise} \end{cases}$$

and

$$W_2 = \begin{cases} W - 70 & \text{if } W \geq 70 \\ 0 & \text{otherwise.} \end{cases}$$

The spline approach worked considerably better than using just a single weather variable. Adding W_1 and W_2 to equation (4.1) reduced the SSR from 0.5411 to 0.5113, a reduction which is significant at the .05 level; the F statistic is 3.613, compared with $F_{.05}(2, 124) = 3.069$. Adding W_1 and W_2 times each of the other variables reduced the SSR to 0.4580, a reduction which is also significant; the F statistic is 2.104, compared with $F_{.05}(10, 116) = 1.913$. Thus the weather variables do appear to add to the specification.

The weather variables alone, however, do not perform nearly as well as OLS. The model

$$C_t = \sum_{j=1}^{5} a_j X_{jt} + b_1 W_{1t} + b_2 W_{2t} + u_t \tag{4.2}$$

had an SSR of 1.7396. Adding the seasonal variables of (4.1) yielded an F statistic of 5.416, compared with $F_{.05}(55, 116) = 1.445$. Thus the seasonally-varying parameters add considerably to the specification based on the weather variables. Using seasonally-varying coefficients alone seems preferable to using (4.2) or (4.3). It would appear that, in this case, the explicit introduction of weather variables cannot replace the seasonally-varying parameters model, but it may be a useful addition.

5. Conclusion

We have argued that the mechanical treatment of seasonality by the use of deseasonalized data or seasonal dummies cannot, in general, be justified on the basis of economic theory. Econometricians should, instead, attempt to build seasonality into their models. If that is not possible, they should at least allow seasonal influences to enter in less restrictive ways than is customary. One way to do so is to allow all coefficients to vary with the season, but OLS estimation of such models is likely to be difficult. An alternative approach involving the use of smoothness priors is described here, and sampling experiments show that it compares well with OLS. The smoothness approach was then applied to estimating the demand for soft drinks, and it performed well compared with the explicit use of weather variables. The smooth seasonality approach thus appears to be practical and potentially useful.

Appendix 1

This appendix shows that $\hat{\boldsymbol{\beta}}$ is unchanged if $k\tilde{\boldsymbol{R}}_1$ in (2.11) is replaced by

$$\bigl((\lambda-1)/\lambda\bigr)^{1/2} k\boldsymbol{R}_1,$$

and the number of zeros in $\tilde{\boldsymbol{Y}}$ is correspondingly increased by one. The only part of (2.10) which is affected by the substitution is $\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}}$, which is originally equal to

$$\boldsymbol{X}^\top\boldsymbol{X} + k^2\boldsymbol{R}_1^{*\top}\boldsymbol{\Omega}^{*\,-1}\boldsymbol{R}_1^* \tag{A.1}$$

and becomes

$$k^2\bigl((\lambda-1)/\lambda\bigr)\boldsymbol{R}_1^\top\boldsymbol{R}_1. \tag{A.2}$$

The inverse of $\boldsymbol{\Omega}^*$ is a matrix with $2(\lambda-1)/\lambda$ on the principal diagonal and $(\lambda-1)/\lambda$ everywhere else. Thus the second term in (A.1) can be written as

$$k^2\bigl((\lambda-1)/\lambda\bigr)\boldsymbol{R}_1^{*\top}\boldsymbol{W}\boldsymbol{R}_1^*, \tag{A.3}$$

where $\boldsymbol{W}$ is a matrix with 2 on the principal diagonal and 1 everywhere else.

Let $\boldsymbol{M}$ be a matrix consisting of an identity matrix of order $\lambda-1$ plus an additional row every element of which is -1. It is easily seen that

$$\boldsymbol{R}_1 = \boldsymbol{M}\boldsymbol{R}_1^*, \tag{A.4}$$

since the identity matrix simply recreates $\boldsymbol{R}_1^*$ and the final row of restores the last row of $\boldsymbol{R}_1$, which is equal to minus the sum of the rows of $\boldsymbol{R}_1^*$. Moreover, it may readily be verified that $\boldsymbol{M}^\top\boldsymbol{M} = \boldsymbol{W}$. Hence

$$\boldsymbol{R}_1^{*\top}\boldsymbol{W}\boldsymbol{R}_1^* = \boldsymbol{R}_1^{*\top}\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{R}_1^* = \boldsymbol{R}_1^\top\boldsymbol{R}_1. \tag{A.5}$$

Thus the second term in (A.1) is equal to the second term in (A.2), which implies that $\hat{\boldsymbol{\beta}}$ is unchanged.

Appendix 2

All data presented are Canadian and, except for the weather variables, they were accessed through the CANSIM data base. The data were published in Table 4 of the original article. They may be accessed at the website of the second author.

CON: volume index of soft drink production, 1961 = 100, monthly, unadjusted.

POP: total noninstitutional population, monthly, unadjusted.

POP25: total noninstitutional population aged 25 and over, monthly, unadjusted.

SDPRICE: consumer price index for soft drinks, monthly, unadjusted.

FPRICE: consumer price index for food, monthly, unadjusted.

CPRICE: consumer price index for all items excluding food, monthly, unadjusted.

QEXP: total personal expenditure on nondurable goods in constant dollars, quarterly, unadjusted.

EXP: monthly interpolation of QEXP. If the month is the middle month of a quarter, EXP is equal to QEXP; otherwise, EXP is equal to two-thirds of QEXP plus one-third of QEXP for the adjacent quarter.

$C = \log(\mathrm{CON}/\mathrm{POP})$.

$Y = \log\bigl((\mathrm{POP} - \mathrm{POP25})/\mathrm{POP}\bigr)$.

$SDP = \log(\mathrm{SDPRICE}/\mathrm{CPRICE})$.

$FP = \log(\mathrm{FPRICE}/\mathrm{CPRICE})$.

$EX = \log(\mathrm{EXP}/\mathrm{POP})$.

WM: average daily maximum temperature in degrees Fahrenheit, Montreal.

WT: average daily maximum temperature in degrees Fahrenheit, Toronto.

WV: average daily maximum temperature in degrees Fahrenheit, Vancouver.

$W = \frac{1}{3}\mathrm{WM} + \frac{1}{2}\mathrm{WT} + \frac{1}{6}\mathrm{WV}$.

References

Jorgenson, Dale W. (1964). "Minimum variance, linear, unbiased seasonal adjustment of economic time series," *Journal of the American Statistical Association*, 59, 681–724.

Lovell, Michael C. (1963). "Seasonal adjustment of economic time series and multiple regression analysis," *Journal of the American Statistical Association*, 58, 993–1010.

Nerlove, Marc (1971). "A note on error component models," *Econometrica*, 39, 383–396.

Poirier, Dale J. (1976). *The Econometrics of Structural Change*, Amsterdam: North-Holland.

Shiller, Robert J. (1973). "A distributed lag estimator derived from smoothness priors," *Econometrica*, 41, 775–789.

Shiskin, Julius, Young, Allan H., and Musgrave, John C. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program*, Bureau of the Census Technical Paper No. 15 (revised), Washington, D.C.: U.S. Department of Commerce.

Sims, Christopher A. (1974). "Seasonality in regression," *Journal of the American Statistical Association*, 69, 618–626.

Taylor, William E. (1974). "Smoothness priors and stochastic prior restrictions in distributed lag estimation," *International Economic Review*, 15, 803–804.

Theil, Henri (1971). *Principles of Econometrics*, New York: John Wiley & Sons.

Table 1. First Experiment: $\sigma = 0.5$

	RMSE			Bias		
Coef.	OLS	Small-k	Large-k	OLS	Small-k	Large-k
a_1	0.9326	0.6411	0.6759	-0.1551	0.0398	0.5185*
a_2	1.1687	0.6637	0.6116	-0.1602	0.0160	0.4310*
a_3	0.8696	0.6244	0.4335	0.0545	-0.0646	-0.0691
a_4	1.1716	0.7285	0.6318	-0.1051	-0.2572*	-0.4533*
a_5	0.8122	0.6119	0.4317	-0.0370	0.0383	-0.0422
a_6	1.1086	0.6855	0.5520	-0.2192	-0.1354	-0.3365*
a_7	0.8375	0.6520	0.8213	-0.1408	-0.2366*	-0.7025*
a_8	0.8932	0.6970	0.7330	-0.0517	-0.1765	-0.5971*
a_9	0.9914	0.6445	0.3956	-0.1440	-0.0540	-0.0711
a_{10}	0.9142	0.6326	0.4862	-0.0019	0.1334	0.3122*
a_{11}	0.8380	0.5489	0.3893	0.0238	-0.0861	-0.1043*
a_{12}	0.8303	0.5892	0.4432	0.1914	0.0709	0.1735*
$\bar{a}$	0.3015	0.2922	0.2915	-0.0621	-0.0593	-0.0784
avg.	0.9473	0.6433	0.5504			
b_1	0.07545	0.05177	0.05191	0.01434*	-0.00137	-0.03841*
b_2	0.09541	0.05449	0.04930	0.01271	-0.00155	-0.03318*
b_3	0.06941	0.05050	0.03590	-0.00597	0.00365	0.00328
b_4	0.08513	0.05237	0.04298	0.00705	0.01773*	0.02791*
b_5	0.06233	0.04685	0.03339	0.00361	-0.00170	0.00918
b_6	0.07783	0.04690	0.03725	0.01503	0.00909	0.02406*
b_7	0.05565	0.04365	0.05332	0.00889	0.01515*	0.04480*
b_8	0.05848	0.04566	0.04785	0.00495	0.01295	0.03879*
b_9	0.06743	0.04561	0.02962	0.00951	0.00369	0.00517
b_{10}	0.06200	0.04247	0.03084	0.00110	-0.00790	-0.01639*
b_{11}	0.05599	0.03702	0.02559	-0.00051	0.00653	0.00361
b_{12}	0.05685	0.04041	0.03215	-0.01498	-0.00674	-0.01441*
avg.	0.06850	0.04648	0.03918			

Table 2. Second Experiment: $\sigma = 1.0$

	RMSE			Bias		
Coef.	OLS	Small-k	Large-k	OLS	Small-k	Large-k
a_1	1.6017	0.9848	1.2296	0.0959	0.2887	1.0634*
a_2	2.1140	1.0642	1.0884	−0.2663	0.2191*	0.8867*
a_3	1.8798	1.0819	0.6567	0.0890	−0.0565	0.1393*
a_4	1.9409	0.9474	0.8047	0.1671	−0.2639*	−0.5174*
a_5	1.5108	0.9078	0.6719	0.0085	0.1083	−0.2819*
a_6	1.9447	1.0204	1.0177	−0.3095	−0.1110	−0.7945*
a_7	1.5287	1.0025	1.4186	−0.0663	−0.3127*	−1.2691*
a_8	1.5233	0.9126	1.2388	0.0568	−0.2110	−1.0719*
a_8	2.2307	1.0511	0.6841	0.0448	0.1358	−0.2623*
a_9	1.7255	1.1010	0.7674	0.0960	0.3437*	0.4245*
a_{10}	1.9102	1.0111	0.6706	0.1146	−0.0854	0.1696
a_{11}	1.8139	0.9894	0.8982	0.1027	0.0616	0.6310*
a_{12}	0.5102	0.5063	0.5074	0.0111	0.0097	−0.0735*
avg.	1.8103	1.0062	0.9289			
b_1	0.13244	0.08351	0.09148	−0.00639	−0.02162*	−0.07681*
b_2	0.17255	0.08866	0.08155	0.02045	−0.01866	−0.06554*
b_3	0.14595	0.08242	0.04850	−0.00845	0.00316	−0.01628*
b_4	0.13823	0.06696	0.05207	−0.01210	0.01781*	0.02538*
b_5	0.11459	0.07022	0.05589	−0.00398	−0.01007	0.03209*
b_6	0.13864	0.07124	0.07009	0.01657	0.00257	0.05518*
b_7	0.09613	0.06281	0.08630	0.00106	0.01698*	0.07609*
b_8	0.09499	0.05835	0.07455	−0.00608	0.01099	0.06246*
b_9	0.14290	0.06802	0.04626	−0.00360	−0.00948	0.01842*
b_{10}	0.11981	0.07612	0.04767	−0.00448	−0.02036	−0.01622*
b_{11}	0.12997	0.06993	0.04960	−0.00291	0.00929	−0.01932*
b_{12}	0.12692	0.06995	0.06681	−0.00632	−0.00357	−0.04705*
$\bar{b}$	0.03535	0.03504	0.03449	−0.00136	−0.00191	0.00237
avg.	0.12943	0.07235	0.06423			

Table 3. Estimates of Soft Drink Demand

Variable	Month	Smoothness	OLS	Lovell	Unified
S	1	−5.19 (7.64)	−6.65 (5.34)	−5.42 (15.64)	
	2	−4.69 (7.10)	−3.45 (2.93)	−5.30 (15.26)	
	3	−4.97 (7.73)	−5.78 (4.96)	−5.40 (15.53)	
	4	−4.81 (7.70)	−4.44 (3.93)	−5.22 (14.99)	
	5	−4.58 (7.46)	−4.58 (4.18)	−5.17 (14.87)	
	6	−4.70 (7.72)	−3.97 (3.75)	−4.98 (14.37)	
	7	−5.07 (7.82)	−6.91 (5.95)	−4.92 (14.29)	
	8	−5.10 (7.59)	−4.12 (3.44)	−4.95 (14.18)	
	9	−5.66 (8.24)	−4.46 (3.51)	−5.16 (14.54)	
	10	−6.16 (8.87)	−8.72 (6.86)	−5.42 (15.04)	
	11	−5.31 (7.62)	−3.70 (2.80)	−5.31 (14.94)	
	12	−5.16 (7.44)	−4.66 (3.68)	−5.25 (14,95)	
	avg.	−5.12	−5.12	−5.21	−2.92 (3.99)
SDP	1	−0.65 (2.64)	−0.72 (2.00)		
	2	−0.08 (0.32)	0.22 (0.58)		
	3	−0.45 (1.62)	−0.52 (1.22)		
	4	−0.63 (2.18)	−0.94 (2.10)		
	5	−0.63 (2.14)	−0.30 (0.65)		
	6	−0.80 (2.69)	−1.10 (2.37)		
	7	−1.01 (3.26)	−0.66 (1.32)		
	8	−1.09 (3.50)	−1.21 (2.43)		
	9	−1.16 (3.71)	−1.48 (2.92)		
	10	−1.07 (3.50)	−0.85 (1.80)		
	11	−0.62 (2.08)	−0.53 (1.07)		
	12	−0.73 (2.83)	−0.88 (2.39)		
	avg.	−0.74	−0.69	−0.69 (5.31)	−0.88 (2.58)
FP	1	0.24 (0.95)	−0.14 (0.26)		
	2	0.14 (0.57)	0.27 (0.54)		
	3	0.08 (0.32)	−0.05 (0.11)		
	4	0.06 (0.24)	−0.11 (0.21)		
	5	0.08 (0.34)	0.49 (1.00)		
	6	0.09 (0.39)	−0.17 (0.37)		
	7	0.19 (0.73)	0.21 (0.36)		
	8	0.29 (1.11)	0.54 (0.93)		
	9	0.36 (1.38)	0.39 (0.71)		
	10	0.40 (1.53)	0.26 (0.46)		
	11	0.37 (1.47)	0.25 (0.48)		
	12	0.34 (1.36)	0.89 (1.68)		
	avg.	0.22	0.23	0.17 (1.14)	1.01 (2.69)

Table continued on next page.

Table 3 (continued). Estimates of Soft Drink Demand

Variable	Month	Smoothness	OLS	Lovell	Unified
EX	1	3.46 (2.31)	6.63 (2.60)		
	2	3.28 (2.20)	0.21 (0.09)		
	3	4.26 (2.98)	5.80 (2.54)		
	4	4.84 (3.52)	5.00 (2.21)		
	5	3.35 (2.36)	2.07 (0.90)		
	6	4.58 (3.25)	4.25 (1.92)		
	7	4.91 (3.18)	7.94 (3.25)		
	8	4.96 (3.33)	3.01 (1.27)		
	9	6.33 (4.39)	4.72 (1.97)		
	10	6.54 (4.79)	10.69 (4.69)		
	11	4.49 (3.17)	1.52 (0.64)		
	12	4.12 (2.81)	3.15 (1.29)		
	avg.	4.59	4.59	4.76 (6.86)	−0.10 (0.08)
Y	1	0.41 (1.25)	−0.22 (0.36)		
	2	0.64 (2.00)	1.09 (1.90)		
	3	0.67 (2.10)	0.32 (0.55)		
	4	0.74 (2.36)	1.05 (1.83)		
	5	0.64 (2.10)	0.45 (0.81)		
	6	0.61 (1.99)	1.12 (2.05)		
	7	0.35 (1.11)	−0.58 (1.00)		
	8	0.37 (1.13)	0.81 (1.34)		
	9	0.34 (1.00)	0.99 (1.54)		
	10	0.19 (0.57)	−1.04 (1.63)		
	11	0.39 (1.13)	1.10 (1.63)		
	12	0.40 (1.20)	0.64 (1.05)		
	avg.	0.48	0.47	0.43 (2.45)	1.42 (3.51)
SSR		0.6422	0.5411	0.7833	5.8110
R^2 (unadjusted)		0.9107	0.9248	0.8911	0.1921
DW Statistic		2.043	1.844	1.963	0.859

Note: The t statistics for the Smoothness estimates are conditional on the prior.